AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Bulletin


ISSN: 0033-2909

2024, Vol. 150, No. 6, 727–766
https://doi.org/10.1037/bul0000430

# Digital Data and Personality: A Systematic Review and Meta-Analysis of Human Perception and Computer Prediction

Joanne Hinds and Adam N. Joinson
Information, Decisions and Operations Division, School of Management, University of Bath

In recent years, our increasing use of technology has resulted in the production of vast amounts of data. Consequently, many researchers have analyzed digital data in attempt to understand its relationship with individuals' personalities. Such endeavors have inspired efforts from divergent fields, resulting in widely dispersed findings that are seldom synthesized. In this two-part study, we draw from two distinct areas of personality prediction across psychology and computer science to explore the convergent validity of self-reports with human perception and machine learning algorithms, the identifiability of the Big Five traits, and the predictability of different types of data. In Study 1, five meta-analyses of human perception studies integrating findings from 24,124 individuals rated across 30 independent samples demonstrated moderate convergent validity across all traits (ranging from $\rho = 0.38$ for Neuroticism, to $\rho = 0.57$ for Openness). In Study 2, a multilevel meta-analysis of computer prediction studies reporting 534 effect sizes across 42 studies also demonstrated moderate convergent validity ($\rho = 0.30$). Multivariate analyses of the significant moderators highlighted that X, Facebook, Sina Weibo, videos, and smartphones had a negative impact on the variance identified. Finally, in synthesizing the extant literature, we discuss the measures used to assess personality and the analytical approaches adopted. We identify the strengths and limitations across each field and explain how interdisciplinary methodologies could advance the testing and development of psychological theory.

***Public Significance Statement***
This systematic review demonstrates that personality traits can be predicted using digital data across both human perception and machine learning algorithms with moderate effects. A further discussion of the types of platforms, digital data, methods, and analyses removes discipline-specific boundaries and reveals some novel future directions that seek to capitalize on new methodologies and advance the development of theory. The implications of these findings are profound across many diverse applications including education, recruitment, marketing, finance, and public health. As such, these findings can inform targeted interventions and communication strategies tailored to individual needs.

*Keywords:* Big Five traits, personality, machine learning, digital footprints, big data

*Supplemental materials:* https://doi.org/10.1037/bul0000430.supp

Our ability to make successful judgments about others' personalities can have major consequences in our lives. Deciding whether someone is trustworthy, reliable, or dangerous, for instance, can impact who we allow to babysit our children, hire for a job, or rent a room to. In recent years, technology has revolutionized the way we communicate and connect with others, and access to such information can profoundly influence our judgments about them. Simultaneously, advances in machine learning and big data analytics have opened new ways to study and derive insights about people's personalities. As such, research has proliferated over the last decade, and work that was traditionally confined to psychology laboratories has now attracted the interests of computer scientists,

This article was published Online First May 16, 2024.
Joanne Hinds https://orcid.org/0000-0001-8960-5361
Adam N. Joinson https://orcid.org/0000-0001-7019-7038
Adam N. Joinson received funding from Grant ESRC Award: ES/N009614/1 from the Centre for Research and Evidence on Security Threats (CREST). The authors have no known conflicts of interest to disclose. Data, analysis code, and research materials are available at https://osf.io/9cndh/.

Joanne Hinds played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing–original draft, and writing–review and editing. Adam N. Joinson played a lead role in funding acquisition and project administration.
Correspondence concerning this article should be addressed to Joanne Hinds, Information, Decisions and Operations Division, School of Management, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom. Email: J.Hinds@bath.ac.uk

marketers, and health practitioners (to name but a few), all keen to predict people's personalities using digital data (e.g., Gladstone et al., 2019; Kosinski et al., 2013; Montag & Elhai, 2019). Consequently, personality research has become highly disparate, with each discipline adopting different conventions, styles, methodologies, and motives. Further, cross-disciplinary communication is rare, meaning that psychologists often lack insight into advanced machine learning techniques, and computer scientists typically lack insight into psychometrics and psychological theory. As our use of devices and generation of digital data continues to grow, bridging this gap is critical for scientific advancement—and for the implementation of personality prediction in business, society, health care, and beyond.

Technological advances also provide access to a broad variety of data (or cues) from which personality can be assessed. For instance, research has demonstrated that personality can be predicted from smartphone logs (Chittaranjan et al., 2013; Stachl, Au, Schoedel, Gosling, et al., 2020), financial transactions (Gladstone et al., 2019), music listening patterns (Anderson et al., 2021), and social media data (e.g., Kosinski et al., 2013; Park et al., 2015; Schwartz et al., 2013). These data reflect individuals' thoughts, feelings, attitudes, and behaviors in real time, and at scale have the potential to provide insights into personality above and beyond the cues that human observers typically rely on when making judgments (i.e., body language, facial expressions, tone of voice, accent, etc.) because they can capture aspects of individuals' personalities that humans may not be able to perceive (e.g., sleeping patterns, physical activity, network ties). However, we currently know little about the relative value of different types of data, and how effective different approaches to assessing personality really are, as most approaches tend to focus on the "accuracy" of human perceptions or computer algorithms and whether predictions from different types of cues are possible. This research therefore seeks to address this gap by answering the following questions: (a) *Convergent validity*—how much do personality assessments based on digital data converge with self-reports (across both human and computer-based approaches)? (b) *Identifiability*—do personality traits differ in their identifiability (both between the different Big Five traits, and across human- and computer-based methods) and (c) *Predictability*—are certain type/s of data (i.e., cues) more indicative of personality than others?

We answer these questions over a series of meta-analyses and discussions, that combine statistical methods alongside narrative syntheses of the findings, methods, and perspectives employed across research on human perception and computer-based prediction. By integrating insights across these two diverse fields, we discuss how each can capitalize on the strengths of the other, and consider the scientific, societal, and ethical implications involved. Finally, we encourage researchers to embrace diversity in future personality research and outline key directions that embrace an interdisciplinary agenda. Over the following sections, we outline the background and underpinnings of both human perception and computer-based personality prediction.

## Origins of Personality Perception Research and the Big Five Traits

Personality is frequently considered to be a set of traits that determine how we think, feel, and behave over time and across different situations (e.g., Funder, 1995, 2012). Although personality is a latent construct that cannot be measured directly, assuming that it is real (Allport, 1937; Funder & West, 1993) and is something that can be discovered has led psychologists to create mechanisms by which personality can be assessed. For instance, in the 1990s, the five-factor model (FFM) of personality (also known as the Big Five, or Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism model) became a well-established framework that researchers use to assess individuals' traits (Digman, 1990; McCrae & Costa, 1999). The FFM is a hierarchical model comprising five basic dimensions–Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Each factor is broad and bipolar (e.g., Extraversion vs. Introversion) and encapsulates several more specific facets (e.g., enthusiasm), that comprise further specific traits (e.g., gregariousness, warmth, excitement-seeking). The FFM suggests that most individual differences can be categorized into these dimensions, and as such remains the most widely used/dominant measure for personality assessment today (as demonstrated across multiple meta-analyses, Connelly & Ones, 2010; Heller et al., 2004; Steel, 2007).

Personality assessments using the FFM have a long history in human personality perception research that date back to the 1970s (e.g., Murray, 1975; Scherer, 1978). Traditional human perception studies involve a *target* (an individual whose personality is assessed), and an *observer* (a person who analyses the target and formulates a judgment on their personality). Observers use "cues" such as body language, facial expressions, accents to formulate their impressions. Such cues help them to gauge what a person is like and will likely influence them in making some form of decision (e.g., whether to lend someone money). In face-to-face interactions, making judgments using these types of cues are relatively straightforward because they are readily available, and observers can monitor how targets react to the conversation/environment around them. Likewise, if individuals are already acquainted, then the observer will be more attuned to the targets' underlying feelings and may recognize if they behave out of character (Biesanz et al., 2007; Colvin & Funder, 1991; Connelly & Ones, 2010; Funder & Colvin, 1988; Funder & West, 1993). Generally, research shows that the more acquainted people are, then the better they can intuit information about each other (Kurtz & Sherker, 2003; Paulhus & Bruce, 1992; Watson et al., 2000). This is because acquainted individuals have access to more information, as they have experienced each other's feelings, attitudes, and behaviors over time and across different contexts. In contrast, strangers (or "zero-acquaintances") lack this privilege; judgments are typically restricted to single instances (e.g., a job interview, an introduction at a party), or even very brief interactions known as "thin slices" (Ambady & Rosenthal, 1992; e.g., a fleeting interaction with a sales assistant). Although zero-acquaintance judgments are based on limited information, research has shown that they can be accurate (Albright et al., 1988; Beer & Watson, 2008; Borkenau & Liebler, 1992; Kenny, 1994).

Online, zero-acquaintance judgments are particularly pertinent because the internet connects us with people over vast distances and provides a mechanism for people to present themselves across different platforms. These judgments can be challenging because targets can self-enhance or misrepresent themselves more easily and frequently than in face-to-face interactions. Online cues are also more open to interpretation—for example, the absence of social contextual cues can mean that a jealous congratulatory expression can be concealed as sincere, or a sarcastic comment can be interpreted as literal.

Using digital traces as cues from which to form personality assessments creates new opportunities to investigate and gain further understanding of personality. Digital traces capture or represent many of the cues we use in offline judgments, but in digitized form (such as language, facial expressions, appearance), as well as "digital manifestations" of a person's behavior (e.g., clicks, number of logins, timestamps). In the context of online environments, digital traces can be used to form impressions of a person's personality in a similar way to how observers use cues in offline judgments. For instance, on social media, individuals can present themselves, however they like, and can assert their personality through *identity claims* (symbolic statements made to declare one's identity, e.g., a framed family photograph, a trophy/certificate, Gosling et al., 2002) such as photos, tweets, usernames, etc. (Hogan, 2010; Krämer & Winter, 2008), although this may be challenged by "friends" who know the person offline as well as online (DeAndrea & Walther, 2011). Observers can utilize these cues to form impressions, for example, they may infer that a person is high in Openness due to their adventurous skydiving photograph.

Individuals also leave *behavioral residue* (traces of behavior, typically left unconsciously, e.g., a few bottles of wine may indicate a party occurred the night before, Gosling et al., 2002) with every action they make, ranging from behaviors that others can see (e.g., a timestamp of a tweet), to actions that may be more difficult for humans to understand or process (e.g., number of text messages sent per day). While maybe not as obvious (or explicitly suggestive of personality) as identity claims, observers may still be able to derive meaning from behavioral residue. A timestamp of a person who tweets late at night may lead an observer to believe they are irresponsible (because they start work early in the morning), for instance. For behavioral residue that is incredibly difficult or near-impossible for humans to quantify (e.g., a person's network structure), computer algorithms can offer a means to interpret relationships between digital traces and personality. Indeed, much of the research in computer science has capitalized on this opportunity by analyzing high-dimensional, fine-grained data aggregated over time and in different contexts.

Taken together, digital data in the form of identity claims and behavioral residue provide both humans and computers with new ways to formulate personality assessments. Still, we know little about how effectively observers and algorithms converge with target reports, and how valuable different types of data are in predicting personality. Before exploring these issues via our meta-analytic investigations, we outline how theories of personality perception have guided research to date and how computational approaches may challenge and extend traditional research practices.

## Theories of Personality Perception

Forming an effective personality impression is complex because personality is a latent construct that cannot be measured directly. As such, for decades researchers have focused on convergent validity or "accuracy" to evaluate how effective personality assessments are and use frameworks such as funder's realistic accuracy model to explore the conditions that make accurate personality judgments possible, including whether cues are relevant, available, detectable, and utilized (e.g., Funder, 1995, 1999). Other theoretical frameworks have also focused on other elements that underpin accurate assessments; these include Kenny's weighted accuracy model (Kenny, 1991) and Kenny's PERSON (personality, error, residual, stereotype, opinion, and norm) model (Kenny, 2004). These include the amount of information they are exposed to, whether they view the same behavior, and whether the target's behavior is consistent. Similarly, Brunswik's lens model (Brunswik, 1956) argues that good judgments are dependent on the extent that observers can distinguish between (and utilize) valid and invalid cues. Although these frameworks were all developed prior to social media, the internet, and digital devices as we know them today, all have formed the basis for many investigations of personality perception in both online and offline environments. As such, research has demonstrated that observers have successfully formed impressions of others by surveying their enunciation (Lippa, 1998), appearance (Naumann et al., 2009), humor (Borkenau et al., 2004), voice quality (Scherer, 1978) as well as their social media profiles (e.g., Back et al., 2010; Darbyshire et al., 2016) and personal websites (Marcus et al., 2006; Vazire & Gosling, 2004).

Although these models readily apply to personality prediction from digital traces, the affordances of computer-mediated communication (CMC) and digital environments can introduce numerous challenges. The absence of social contextual cues, the opportunity to present oneself more favorably, and the increased likelihood of performing zero-acquaintance judgments (or at least not knowing targets well/offline) is likely to make achieving accurate judgments more laborious. For instance, a well-meaning tweet will not be considered friendly if the tone is hostile (relevance), a posed photograph of a couple may not reveal a strained or unhappy relationship (availability), and a solemn post may be interpreted as attention-seeking, rather than a cry for help (utilization). Researchers sometimes attempt to disentangle these concepts by using Brunswik lens model analyses. According to Brunswik's lens model (Brunswik, 1956), observers can use elements in the environment (i.e., digital traces) as a lens to indirectly perceive underlying constructs. For example, an observer may infer that a target is an extravert after seeing how many friends they have (both in comments in their posts and their photos). For their judgment to be accurate, the observer must use the cues available (cue utilization), and in turn these cues must be valid (cue validity), (i.e., they must be associated with the underlying personality trait). For instance, Back et al. (2008) found that observers utilized the number of full stops and characters in targets' email addresses to infer Conscientiousness, and Hall et al. (2014) found that observers utilized emoticon use and use of positive affect in status updates to infer Extraversion. In the former case, characters and full stops were not found to be valid indicators of Conscientiousness, whereas in the latter, both were utilized by observers and were valid indicators of Extraversion.

The possibility of using and analyzing cues in this way creates some new opportunities for predicting personality with computer algorithms. For instance, algorithms can "utilize" cues in aggregated combinations to potentially discover relations between personality and digital traces that humans could not perceive. Likewise, the ability to analyze masses of data (at population level) enables researchers to gain an understanding of personality in ecologically valid settings. It could also enable them to predict personality from individuals' behavior (i.e., patterns of behavior recorded without a person thinking or actively trying to control how they may appear, such as their Global Positioning System (GPS) movement patterns, or smartphone usage data). This may help to overcome some of the measurement issues relating to self-reports where estimates of behavior have often failed to map to real life (e.g., Andrews et al., 2015; Davidson et al., 2021; Parry et al., 2021). In the following section, we explore these properties further by discussing key methodological differences between human perception studies in psychology and machine learning studies in computer science.

## Computer-Based Personality Prediction

Individuals' constant use of the internet, digital devices, and the internet of things has led to a surge in research using computer-based prediction (i.e., machine learning) to predict personality using digital data. Computer-based personality research has become so popular that numerous computer science conferences have dedicated annual competitions and workshops that seek to experiment with different data sets and improve algorithms' predictive performance (e.g., the myPersonality data set[1] and the Plagiarism Analysis, Author Identification, and Near-Duplicate Detection at the Conference and Labs of the Evaluation Forum). Together, this work has demonstrated that a wide variety of data can be used effectively including metadata (e.g., number of friends, photos, tweets, Bachrach et al., 2012; Ngatirin et al., 2016; Skowron et al., 2016; network data, Chapsky, 2011; de Oliveira et al., 2011; smartphone logs, Chittaranjan et al., 2011, 2013; Kambham et al., 2018; Yakoub et al., 2015; and language, Nowson et al., 2007; Park et al., 2015; Schwartz et al., 2013; Shen et al., 2013). Computational approaches have also started to explore whether facets (specific aspects of each trait) and nuances (item-specific aspects of traits below facets, Mõttus et al., 2017; can be predicted using digital data). For example, Stachl, Au, Schoedel, Gosling, et al. (2020) predicted facets using behavioral data from smartphones (e.g., music consumption, app usage, mobility), and Hall and Matz (2020) found that nuance-based approaches were better at predicting personality than trait-based approaches.

The process of computer-based personality prediction usually involves training an algorithm to predict individuals' personalities based on a set of cues or variables (known as "features" in machine learning) and their self-reported personality surveys. Typically, the cues are partitioned into a "training" set and a "testing" set. The algorithm learns to predict personality from the training set, and its performance is evaluated using the testing set (known as "cross validation")—see Bleidorn and Hopwood (2019), and Stachl, Pargent, et al. (2020) for detailed overviews on this process. Computer-based personality prediction readily applies to many of the theoretical concepts outlined previously, in terms of how algorithms work, and with respect to the types of cues/data they can process. For instance, algorithms can be programmed to predict personality from digital data in the form of identity claims and behavioral residue (e.g., photos, status updates, and number of likes, groups, etc., respectively). Algorithms can analyze both types of digital traces at mass scale, to fine granularities, and in aggregate. That is, sample sizes can comprise thousands or millions of features, digital traces can be measured in minute quantities (e.g., nanoseconds or milliseconds), features can be combined to discover patterns that may not be visible when analyzed independently, and data from one platform (e.g., Facebook) can be used to train an algorithm to predict personality from another platform (e.g., Instagram), a process known as *cross media learning* (e.g., Peng et al., 2017).

These capabilities provide researchers with opportunities to test psychological theories with cues that humans could not mentally process or even comprehend when making personality predictions. This is particularly true for behavioral residue, because such cues represent actions that a target makes (e.g., clicks), or objective information about their identity (e.g., number of friends) rather than how a target appears or presents themselves online. The ability to harvest and analyze behavioral residue also enables researchers to pair personality measures with behavior (or data gained from participants' behavior "in the wild"), a methodology that is typically underrepresented yet repeatedly called for in personality research (Doliński, 2018; Funder, 1995; Hinds & Joinson, 2019).

Computer scientists tend to focus on optimizing the performance of algorithms, typically through meticulous rounds of data cleaning, processing, and testing. This process of refining and optimizing an algorithm's performance has some conceptual similarities with Funder's RAM as they must be able to *utilize* information that is *relevant, available*, and *detectable*. To optimize cue utilization, relevant cues must be detected, and irrelevant cues must be removed (a process known as "feature selection"). Then redundant cues are combined ("dimension reduction"), and each cue is weighted to improve its use in the following prediction (Park et al., 2015). Likewise, the concepts of *cue validity* and *cue utilization* in Brunswik's lens model also apply to computer prediction in a similar way (i.e., digital traces must *validly* represent the target's personality, and the algorithm must be able to *utilize* them effectively). Nonetheless, in contrast to human perception, the features used are predetermined by the researchers when they design the study.[2] This means that researchers can experiment with different features (or combinations of features) without any preconceptions of how they may link to personality (i.e., their validity). Algorithms can work "bottom-up" meaning they can potentially discover unusual or unanticipated patterns in the data—but the extent that these cues are utilized when doing so are determined by the researchers.

The process of optimizing an algorithm's performance is markedly different to the way that human perception studies tend to be designed and conducted, as psychologists do not tend to train observers or incentivize them to perform good judgments. Instead, observers are asked to formulate a judgment based on visible cues to ascertain whether observers' ratings converge with targets' self-reports. Moreover, psychologists tend to be concerned with explaining the relationship between digital data and personality, usually through the lens of a theoretical framework (such a Funder's RAM or Brunswick's lens model), whereas computer scientists are predominantly focused on data-driven predictions of personality (Hinds & Joinson, 2019; Yarkoni & Westfall, 2017).

[1] myPersonality was a Facebook application created by David Stillwell and Michal Kosinski in 2007. It provided users with personality profiles after they completed a questionnaire and collected their Facebook data including likes, status updates, demographic information, etc. Over 5 years, the data set grew to comprise data from over 6 million users and many workshops and reports have published findings relating to what traits and other demographic attributes could be inferred from Facebook data, see: https://mypersonality.org.

[2] In human perception studies, researchers may predetermine the cues/features to some extent, that is, they may provide observers with specific cues to base judgments on (e.g., email addresses, profile pictures, likes, rather than whole social media profiles). Subsequent Brunswik lens analyses of elements comprising those cues may be performed to ascertain which elements are valid and utilized (e.g., number of underscores, or.com in email addresses Back et al., 2008). Yet, unlike computer prediction studies, such information is not fed back into the process of training observers or used to improve observer ratings.

Although the perspectives, motivations, and methodologies across both disciplines are vastly different, neither are perfect, and the strengths of one have the potential to improve the weaknesses of the other. For example, computational approaches could further understanding of the relationship between different types of data and personality by integrating psychological theories into the process of experimenting with and testing out different types of data. Psychological approaches could utilize insights gained from data-driven methods to improve how observers use digital cues and to inform how personality is conceptualized and assessed. As a first step toward building a bridge across these two disciplines, we therefore seek to evaluate how effective current approaches are. We adopt the stance that different methods have the potential to access different kinds of valid information about personality (e.g., automated analyses of a person's smartphone Bluetooth usage may offer different insights to an observer perusing an Instagram profile) and assume that no method is more valid than any other a priori. By assessing the *convergent validity* of human perception and computer prediction with target reports, we compare the effectiveness of both approaches and explore the relative value of different types of data. Specifically, we seek to answer the following: (a) Convergent validity—how much do personality assessments based on digital data converge with self-reports? (b) Identifiability—do personality traits differ in their identifiability (both between the different Big Five traits, and across human- and computer-based methods), and (c) Predictability—are certain types of data (i.e., cues) more indicative of personality than others?

In investigating these questions, we reflect on how different types of digital data and methodological approaches could inform new interdisciplinary perspectives. As such, we outline key directions for future research and consider the broader implications of personality assessments for individuals, organizations, and society. We argue that such insights will be critical to progressing scientific work over the coming years, particularly as our interactions with technology, digital data, and each other become increasingly intertwined.

## The Present Study

Our review constitutes a unique "hybrid" systematic review and meta-analytic approach, whereby a series of meta-analyses (of both human perception and computer-based personality prediction) are presented alongside a narrative review of existing methodologies, and findings. We adopted this methodology so that we could exhaustively synthesize the diverse approaches across both disciplines and retain research that may otherwise be excluded through the constraints imposed in meta-analytic investigations. For instance, many computer prediction studies do not report effect sizes, or statistics that are convertible to effect sizes. Instead, studies often focus on reporting error rates (e.g., mean absolute error, root-mean-square error) or other metrics related to the performance of algorithms (e.g., accuracy, f-score). Such studies would need to be excluded from a meta-analysis despite meeting the broader inclusion criteria. Thus, only reporting meta-analytic analyses would bias our evidence base in favor of effect size reporting, meaning we would lose the many valuable insights, trends, methodologies, and findings presented in the other reports. This was also inspired by Slavin's work on "best evidence synthesis", that argues that the "best evidence" may arise outside of reported effect sizes, and that they should be included in the review as an adjunct to the literature meeting the inclusion criteria. Similar recommendations have also been proposed by Jackson and Waters (2005) and M. Campbell et al. (2020).

Moreover, although there are several related reviews on personality perception/prediction, existing work tends to be discipline specific (i.e., it focuses on machine learning or human perception only, e.g., Connelly & Ones, 2010; Tskhay & Rule, 2014), adopts a narrower focus on digital data, for instance, by focusing on social media data only (e.g., Azucar et al., 2018; Marengo & Montag, 2020) and dedicates little attention to theoretical insight. As such, our research is novel in that it removes discipline-specific boundaries, it meta-analyses effect sizes and reviews the current state-of-the-art across all types of digital data, platforms, and devices. Accompanying our narrative discussion of approaches, methodologies, and findings, we present a taxonomy of digital data used to predict personality in computer prediction studies (see Table 1). The taxonomy shows example references from our data set, encompassing reports that were included in the meta-analysis, as well as those omitted due to their metric reporting. This approach offers a comprehensive overview of the research landscape, capturing detail from all studies that predicted personality using digital data. Further, our Supplemental Materials present descriptive statistics of the data collected and synthesized, and all data relating to the human and computer prediction studies extracted from our searches (see Supplemental Files 1, 2, and 3, respectively).

# Method

To synthesize existing evidence on human- and computer-based personality predictions, we performed an extensive literature search that informed the two separate studies. Study 1 reports five meta-analytic investigations for human perception, and Study 2 reports a multilevel meta-analysis for computer-based personality prediction. Across both studies, our extraction of data and subsequent discussion followed the Preferred Reporting of Items for Systematic Reviews and Meta-Analysis guidelines (Page et al., 2021) and the APA Meta-Analysis Reporting Standards (American Psychological Association, 2010). Further, to capture data in accordance with the different aims, methodologies, and reporting conventions across both studies, we created separate inclusion/exclusion criteria for human perception and computer-based prediction. Details on the meta-analytic approaches, search procedure, coding of study variables, and inclusion/exclusion are described below.

## Meta-Analytic Approaches

We selected Pearson's $r$ as our effect size metric because it is the most comparable measure of convergent validity used across both human- and computer-based prediction studies. In human perception studies, self-other agreement is typically measured by calculating the correlation between observer and target personality scores. A positive value indicates a direct relationship—meaning as one score increases, the other also increases—while a negative value suggests an inverse relationship. For computer-based predictions, studies report a variety of statistics relating to an algorithm's performance (e.g., root-mean-square error, accuracy, precision, recall) as well as (or instead of) correlations between an algorithm's score and target's personality score. In this context, the algorithm assumes the role of the "other" in self-other agreement between computer prediction and a target's personality rating. Akin to self-

**Table 1**
*Taxonomy of Digital Data (Features) Used to Predict Personality in the Computer Prediction Studies*

| Type of data | Platform | Features used to predict personality | Reference |
|---|---|---|---|
| Language | YouTube, Github, Buzz Metric, Facebook, X, Sina Weibo, Gmail, World of Warcraft | Ngrams, bigrams, number of words used in different categories (e.g., family, occupation, affect), sentiment (e.g., positive/negative words) | Alam and Riccardi (2014), Bayot et al. (2015), Giménez et al. (2015), Hickman et al. (2022), Philip et al. (2019), van Mil (2020), and Yakoub et al. (2015) |
| Metadata | Facebook, X, Outlook, Gmail, LiveJournal, Instagram, Sina Weibo, personal computer | Number of @mentions, replies, hashtags, verbs, nouns, exclamation marks, URLs, swear words | Bayot et al. (2015), Grover and Mark (2017), Sarkar et al. (2014), and Titov et al. (2019) |
| Demographics | YouTube, Facebook, Sina Weibo, Spotify | Name, birthday, relationship status, religion, education, gender | Anderson et al. (2021), Chapsky (2011), Farnadi et al. (2014), and Golbeck et al. (2011) |
| Network data | XING, Facebook, X, World of Warcraft, smartphones | Network size, betweenness, density, brokerage, transitivity | de Oliveira et al. (2011), Golbeck (2016), and Kafeza et al. (2014) |
| Call logs and messages | Smartphones | Number of outgoing calls, average duration of outgoing calls, unique contacts called, time of calls | Chittaranjan et al. (2011, 2013), de Oliveira et al. (2011), and Stachl, Pargent, et al. (2020) |
| Application usage | Smartphones | Number of uses of camera app, chat app, installs per month, app updates | Chittaranjan et al. (2011, 2013) and Xu et al. (2016) |
| Bluetooth data | Smartphones | Number of unique Bluetooth IDs, seen more than 4, 9, 19 slots | Chittaranjan et al. (2011, 2013) and Rüegger et al. (2020) |
| Location data | Smartphones, beacons | Radius of gyration, distance traveled per day, no. places where phone calls are made, Global Positioning System data, work activities | De Montjoye et al. (2013), Kambham et al. (2018), Lepri et al. (2016), Robles-Granda et al. (2021), and Wang, Harari, et al. (2018) |
| Physical activity | Accelerometers | Physical activity intensity in daytime, evening, acceleration (Hz) during weekends | Gao et al. (2019) and Rüegger et al. (2020) |
| Microphone | Smartphone | Average amplitude (dB), voice data | Rüegger et al. (2020) and Wang, Harari, et al. (2018) |
| Audio | YouTube, video-based interviews | Speaking time, average length of speaking segments, spectral entropy, energy, pitch | Alam and Riccardi (2014), Aydin et al. (2016), Biel and Gatica-Perez (2014), Gorbova et al. (2017), and Hickman et al. (2022) |
| Facial expressions and gestures | YouTube, video-based interviews | Anger, contempt, smile, looking time, number of looking turns, proximity to the camera | Alam and Riccardi (2014), Farnadi et al. (2014, 2016), Gorbova et al. (2017), and Hickman et al. (2022) |
| Multimodal | YouTube | Looking while speaking, looking while not speaking | Biel and Gatica-Perez (2013, 2014) |
| Behavioral data | Online learning environment | Data relating to attendance, course viewing, quiz results, tests, exams | Lai et al. (2020) |
| Photos | X, Flickr, Instagram, Sina Weibo, Facebook | 5 o'clock shadow, grey hair, heavy makeup, average rule of thirds, color, brightness, saturation | Celli et al. (2014), Dhall and Hoey (2016), Ferwerda and Tkalcic (2018), Torfason et al. (2016), and Xiong et al. (2016) |
| Moods | Spotify | Defiant, brooding, rowdy, sentimental | Anderson et al. (2021) |
| Genres | Spotify | Reggaeton, folk, death metal, rock | Anderson et al. (2021) |
| Behavioral data | Spotify | Streamed from tablet, average daily skip rate, % streams with low acoustic value | Anderson et al. (2021) |
| Physiological data | Wearable sensors | Heart rate, stress, step count, calories burned, rapid eye movement sleep | Robles-Granda et al. (2021) |
| Behavioral data | Age of Empire II, StarCraft | Food collected, gold collected, total castles, map explored, avatar ID | Halim et al. (2017) |
| Transaction data | U.K.-based money management application | Bond income, council tax, eye care, flights, holiday, pension, toys | Gladstone et al. (2019) |
| Likes | Facebook | Status updates, comments, pages, musical artists | Kosinski et al. (2013), Nave et al. (2018) and Youyou et al. (2015) |
| Interests | Facebook | Movie genres, music genres, favorite tv shows, books, quotes | Chapsky (2011) and Golbeck et al. (2011) |
| Privacy settings | Sina Weibo | Blocking private messages/comments sent by strangers | Li et al. (2014) |
| Motion, posture | Smart home environment | Motion, posture | Dotti et al. (2018) |

other agreement in human perception, a positive correlation suggests that the algorithm's predictions align closely with the target's scores, and a negative correlation implies divergence. Given that many of the statistics reported in machine learning studies can be converted to correlations, Pearson's $r$ therefore seemed the most appropriate effect size to use in our meta-analyses. For a detailed explanation of converting these metrics to correlations, refer to Study 2: Calculation of Effect Sizes–Computer Prediction.

In Study 1, we performed five Schmidt–Hunter meta-analyses (one for each trait; Schmidt & Hunter, 2015), and in Study 2, we adopted a multilevel meta-analytic approach. Each of these approaches were best suited to the corresponding data; the Schmidt and Hunter (2015) procedure accounts for the effects of study artifacts (e.g., sampling error, measurement error), enabling analyses to reflect the correlations more effectively among constructs. Alternatively, a large amount of the data in the computer prediction studies were nonindependent (due to researchers reanalyzing the same data sets across many of the studies), thus a multilevel approach enables better handling of the data. Further detail on these approaches and the associated decisions underlying them can be found in the respective sections on Calculating Effect Sizes, and Multilevel Meta-Analysis.

## Literature Search Procedure

To capture the breadth of work in this area, our search included journal articles, conference articles, and grey literature (including unpublished dissertations, and theses). To identify relevant reports, we used five systematic search strategies. First, we performed searches in the Web of Science, the Institute of Electrical and Electronics Engineers (IEEE), and the Association of Computing Machinery (ACM) online libraries for all relevant reports by searching for personality and topic-related terms. Because this area of research has proliferated across multiple disciplines and subtopics (each with different styles and reporting conventions), we used a broad variety of search terms in attempting to locate reports. Our search terms were applied to each database as follows: Web of Science—search by topic (predict or identify or detect* or Facebook or Twitter[3] or Instagram or YouTube) and (person* or individual difference) and (digital or internet or online or computer-mediated) and (social* or big 5 or dark triad or priva* or web* or mobile* or sms or shop* or date or dating or query or language or achievement or affiliation or self-esteem or intelligen* or intelligence quotient or leader* or self-affirm* or influen* or big data or extraversion or neuroticism or openness or conscientiousness or agreeableness or narcissis* or pychopath* or sociopath* or decept* or deceiv*); IEEE—predict* personality (search with all metadata); ACM—search where title matches all "predict personality" and where abstract matches all "predict personality."

Across each of the databases, we searched for reports published up until the start of January 2022 (i.e., our search had a cut-off date of January 5, 2022). The breadth of this search strategy generated 33,193 reports in total (Web of Science = 22,594, IEEE = 538, and ACM = 10,061). The Web of Science search generated a substantial number of reports that were extraneous to our criteria, that is, they were situated in different fields (e.g., physics, engineering), or were exploring subject matter that was obviously redundant (e.g., psychiatry). Such reports were excluded from our search by filtering out reports according to category (i.e., all fields/subject categories that were off topic) and were removed. This procedure reduced the set of reports extracted from Web of Science to 13,599. Next, reports obtained via all three database searches were assessed for potential eligibility by screening their titles and abstracts. Figure 1 displays a flowchart with a detailed breakdown of the number of reports included/excluded at each stage of the systematic search process.

Second, from the 403 set of reports retrieved, we closely scrutinized abstracts and detailed information in their methods, procedures, etc. to determine eligibility. Third, from the set of reports that met our inclusion criteria (detailed in the corresponding sections below), we examined the references for relevant reports and retrieved and screened all further reports that also met these criteria. This step was performed iteratively on each report added to the set until no further reports were retrieved. In other words, we inspected the references of each relevant report (to retrieve further reports) and stopped when no further reports met our criteria. Additionally, we applied this procedure to all relevant references reported in prior meta-analyses and reviews (Azucar et al., 2018; Connelly & Ones, 2010; Kedar & Bormane, 2015; Marengo & Montag, 2020; Settanni et al., 2018; Tskhay & Rule, 2014; Vinciarelli & Mohammadi, 2014).

Fourth, we hand searched Google Scholar pages, ResearchGate profiles, and academic/personal websites of experts in the area, as well as all first authors from the set of reports retrieved. Finally, we contacted all corresponding authors from the reports in our set and inquired if they had any unpublished data or reports that we had not yet located. For all additional reports retrieved, we iteratively examined their references until no further reports were identified (as per Step 3).

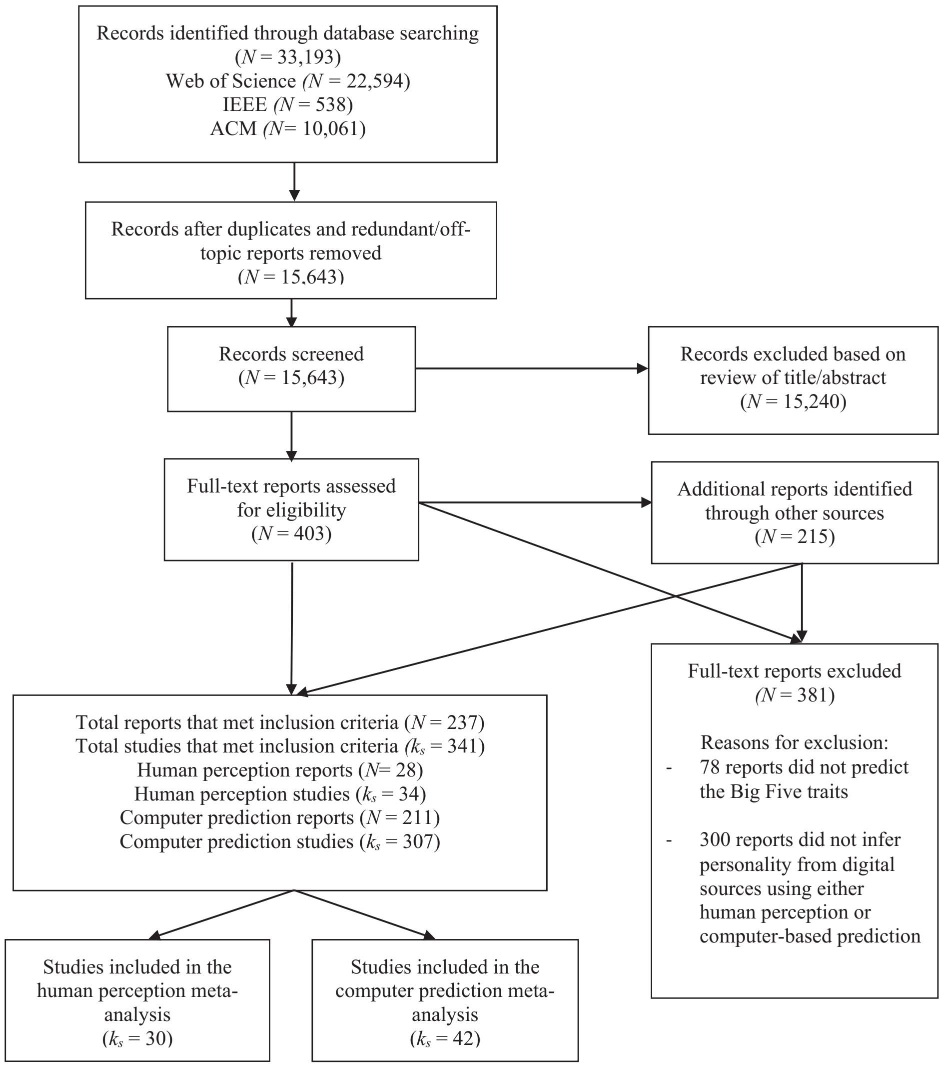
This procedure yielded 237 reports and 341 studies that met our inclusion criteria. These reports were subsequently partitioned into human perception ($N$ = 28 reports, $k_s$ = 34 studies) and computer prediction ($N$ = 211 reports, $k_s$ = 307 studies). All these studies formed the basis for our discussion of theoretical applications, methodologies, and future directions (see the General Discussion). Then, effect sizes were extracted (and transformed where necessary) for inclusion in the meta-analyses. This process resulted in a total of 30 effect sizes that were included in the human perception studies and 534 effect sizes from 42 studies of computer-based personality prediction.

## Coding of Study Variables

We developed a comprehensive codebook for recording study characteristics, effect sizes, and statistics. Specifically, this included: reference information (title, authors, publication year), sample sizes (number of observers/targets in human perception, and number of targets in computer-based prediction), platform/device (e.g. Facebook, X, YouTube, smartphone), type of digital footprint studied (e.g. status updates, number of friends, number of photos uploaded), type of personality questionnaire used (including name and number of questions), effect size (e.g., Pearson's $r$, Kendall's $\tau$, $R^2$), and statistical information relating to the predictive performance of

[3] The term "Twitter" was included in our search because the social media site now known as "X" changed name in 2023 after the search was performed. As such, all reports retrieved from our searches referred to the platform as Twitter. That said, throughout this article, we use the term "X" when discussing the platform to align with current conventions at the time of writing.

**Figure 1**
*PRISMA Flowchart Summarizing Database Search and Report Screening*




*Note*. PRISMA = Preferred Reporting of Items for Systematic Reviews and Meta-Analysis; IEEE = Institute of Electrical and Electronics Engineers; ACM = Association for Computing Machinery.

computer algorithms (e.g., Mean Absolute Error, Root-Mean-Square Error, Precision, Recall, *F*-Score, Area Under the ROC Curve).

We also coded several variables that were specific to either human perception or computer-based studies. For human perception studies, this included the type of questionnaire and number of questions completed by observers. For computer prediction studies, this included the type of machine learning algorithm used for prediction (e.g., Support Vector Machines, Random Forests, Neural Networks). All studies were coded by the first author (a researcher experienced in systematic reviews and interested in personality and meta-analysis) and 50% of randomly selected studies were second coded by a second coder (a research assistant with experience in systematic reviews and meta-analysis) to assess reliability of study coding. Cohen's κ was used to assess interrater agreement, and demonstrated high levels of consensus, $\kappa$ = 0.97 (95%CI [.94, 1.00]). All discrepancies that occurred in data extraction were resolved through discussion. An overview of all studies coded are provided in Supplemental Files 2 and 3. Descriptive statistics providing an overview of the studies analyses are provided in Appendix B (Supplemental File 1).

## Inclusion and Exclusion Criteria

### *Study 1: Human Perception*

Our inclusion criteria for the human perception reports included: (a) studies had to examine personality perception where observers formed judgments by viewing some form of digital footprint (e.g., Facebook profile, tweets, username); (b) studies had to report findings from at least two samples, that included observers and targets. The digital footprint had to be created/produced by the target and the target had to have completed a self-report questionnaire estimating at least one personality factor. Similarly, the observers had to have also completed a personality assessment (either the same or an alternative version) in order to rate the target's personality. Observer questionnaires could be completed by zero-acquaintances or informants; (c) studies had to have reported a measure of agreement about targets' personalities, in the form of self-other agreement (typically referred to as accuracy). In most cases, this statistic was denoted by Pearson's *r*, that represented the aggregate ratings between observers and targets.

### *Study 2: Computer-Based Prediction*

Our inclusion criteria for the computer-based reports included: (a) studies had to include an examination of personality prediction/detection, where machine learning algorithms were trained to detect personality traits from a set of training data. The training data had to be derived from some form of digital footprint; (b) studies had to report findings where the algorithm was trained on data (derived from a target's digital footprints); (c) studies had to report a measure of the algorithm's performance. These criteria were used to select studies that were included in the overall broader review, however, for inclusion in the meta-analysis, we included one additional criterion: (d) studies had to report findings where an algorithm's performance was correlated with targets' self-reported personality scores (or where a measure of the algorithm's performance could be converted to Pearson's *r*).

Similar to human perception, we sought out assessments of convergent validity in the form of Pearson's *r* correlations between the algorithm's personality prediction and targets' personality score. Although psychologists tend to focus on analysis/reporting effect size, computer scientists usually focus on evaluating an algorithm's performance by measuring error rather than convergent validity (i.e., root-mean-square error, mean absolute error, F1 score, precision, recall, and so forth). Often such metrics cannot be converted to effect sizes, or the detail required to do so is not reported. Thus, the inclusion of the additional criterion (d) was vital to ensure that insights could be gained from all studies, irrespective of their inclusion in the meta-analysis. In turn, encompassing statistical analyses and a narrative synthesis of existing evidence enables a richer dialogue that considers key trends, findings, and future directions. As such, insights obtained from studies reporting effect sizes and other performance metrics are presented in our taxonomy (see Table 1) and in the General Discussion.

## Calculation of Effect Sizes

### *Study 1: Human Perception*

In total, there were 24 reports that met our inclusion criteria. Within this set, 30 separate studies reported effect sizes across all of the Big Five personality traits (or a subset thereof). Specifically, the breakdown for the number of studies across each trait included: $k_s = 29$ (Openness); $k_s = 29$ (Conscientiousness); $k_s = 30$ (Extraversion); $k_s = 29$ (Agreeableness); $k_s = 28$ (Neuroticism). Although there was variance across reports in terms of the methods, measures, and reporting of statistics, most studies ($k_s = 33$) reported self-other agreement (i.e., effect sizes) as Pearson's *r* correlations. One study reported self-other agreement as a Kendall's *W* correlation that was converted to Pearson's *r* using Gilpin (1993).

Five meta-analyses were performed (one for each trait) using the Schmidt–Hunter approach (Schmidt & Hunter, 2015), that offers numerous advantages over other meta-analytic processes for this type of research and is commonly used in meta-analyses in industrial-organizational psychology. The Schmidt–Hunter procedure accounts for the effects of study artifacts (e.g., sampling error, measurement error) that introduce variability across correlations that are not attributable to true variability in population correlations. Therefore, by correcting these statistical artifacts, Schmidt–Hunter meta-analyses can more accurately reflect the correlations among constructs. We performed a random effects meta-analysis using Schmidt and Hunter's (2015) psychometric approach. Random effects models (in contrast to fixed-effects models) do not assume that all studies are from an identical population, or that subjects were tested under similar conditions (Schmidt et al., 2009). Thus, random effects models account for study heterogeneity and more accurately estimate confidence intervals (CIs) around point estimates that helps to avoid Type I errors when identifying moderators (Hunter & Schmidt, 2000; Schmidt et al., 2009).

To perform the analyses, we used individual correction methods to correct for artifacts (where each correlation is corrected individually, then the meta-analysis is performed on the corrected correlations). In making adjustments for measurement error, we primarily needed to account for the different methodological and reporting approaches employed to ensure that effect sizes were standardized—this process is outlined as follows. First, across the set of studies, the number of observers and targets recruited varied substantially, that is, the number of observers included ranged from three to 17,662, and the number of targets ranged from 25 to 17,662. As such, studies employed a variety of approaches to formulate observer ratings when assessing self-other agreement. These included: (a) creating groups of observers that rated a subset of the target pool ($k_s = 8$; e.g., see, Back et al., 2008; Hickman et al., 2019, Sample 2), (b) instructing each observer to rate one target, such that one observer rated one target ($k_s = 4$; e.g., see, Darbyshire et al., 2016; Park et al., 2015), and (c) instructing all observers to rate all targets ($k_s = 20$; e.g., see, Gosling et al., 2007; Kluemper et al., 2012; Qiu et al., 2012). Similarly, studies also varied in their calculations of self-other agreement, with some reports presenting a measure of a single observer's rating correlated with targets' self-reports (e.g., Darbyshire et al., 2016; Gosling et al., 2007) and others reporting aggregated multiple observer ratings with targets' self-reports (Hall et al., 2014; Hickman et al., 2019; Stopfer et al., 2014)[4]. In these instances, the number of observers rating each target varied considerably across the studies, for example, in Back et al. (2008) groups of 24–26 observers rated subsets of targets (i.e., there were 599 targets in total, and each observer rated 150 targets) whereas in Hickman et al. (2019), three observers rated all 177 targets.

Typically, interrater reliabilities for observer reports are relatively low, especially when observers are unacquainted with the targets (e.g., Connelly & Chang, 2016; Norman & Goldberg, 1966). Aggregating the ratings of multiple observers can therefore have a substantial impact on self-other agreement, because they generate higher interrater reliabilities, that in turn increases the self-other

[4] Some studies also reported both single and aggregate observer-self ratings. In these cases, the effect size reported from a single observer was included in the meta-analysis.

correlation. For instance, in Back et al.'s (2010) study, self-other agreement for Openness was $r = 0.24$ for a single observer, but $r = 0.41$ when 9 or 10 observer ratings were aggregated.

From a meta-analytic perspective, the heterogeneity in the number of observers used in each study is problematic because studies that only report effect sizes from large numbers of aggregated observers will therefore increase the overall effect size that could cause issues with subsequent moderator analyses. To address this, we made several adjustments to the effect sizes, so that the number of observers were held constant across each study. First, for studies that only reported aggregated observer ratings, we estimated the effect size that would have resulted from a single observer by disattenuating the correlations based on a multiother composite for interrater unreliability for $k$ observers (Connelly & Ones, 2010) along with the internal reliabilities (if reported) or test–retest reliabilities (imputed from the corresponding personality scale used, when internal reliabilities were not reported) for target reports. Where studies reported other-rater interrater reliability (i.e., consensus in the form of intraclass correlations [ICC]), we used the ICC to estimate the effect size for a single-other pair. Some studies reported ICC at the level of a single rater [i.e., ICC(2, 1)], thus in these instances, the ICC was used to estimate the correlation of a single-other pair accordingly. For studies that only reported ICC from multiple observers [i.e., ICC(2, $k$)], we used the Spearman–Brown formula to estimate the interrater reliability of $k$ raters. These results were then used to estimate the single-rater correlations across the corresponding multiother composites.

In some instances, the studies included did not report enough information to perform these calculations. We therefore adopted several approaches, outlined as follows. First, for studies that did not report ICCs, we selected interrater reliability values from another study that most closely matched the effect size on the Big Five trait measured, the number of observers, and the type of data used (similar to Connelly & Ones, 2010 approach in Study 2). Second, some studies did not disclose whether self-other agreement was aggregated or reported at the level of a single observer. In these circumstances, we adopted a conservative approach and assumed self-other agreement was aggregated. Then, we followed the above process for disattenuating the correlations with ICCs. Third, some studies reported self-other agreement that was calculated by merging multiple scales (e.g., Back et al., 2010; Park et al., 2015). Here, we also adopted a conservative approach and disattenuated the correlations using the lowest corresponding reliability. Fourth, where groups of observers were used to rate subsets of targets and the size of the observer groups varied in size (i.e., Hickman et al., 2022, Study 2 used groups that comprised four—seven observers, and no further detail was reported), we disattenuated the correlations based on the overall average group size.

There was no range restriction across the set of studies, so no corrections were made in this respect. Further, effect sizes were included in the meta-analysis as Pearson's $r$ correlations and were not transformed to Fisher's $z$ scores (as recommended by some meta-analytic procedures e.g., Borenstein et al., 2009). Schmidt and Hunter (2015) recommend meta-analyzing untransformed correlations because Fisher's $z$ transformation yields an upward bias when estimating the mean correlation that is typically higher than the bias resulting from untransformed correlations. The meta-analyses were conducted in R (Version 4.2.1) using the *psychmeta* package (Dahlke & Wiernik, 2019).

### *Sensitivity Analyses and Publication Bias*

We performed two types of sensitivity analyses to explore the effect of outliers on effect sizes estimates that included winsorizing large and small sample sizes, and a sensitivity analysis where an outlier was removed. Most studies within our set ranged in sample size, from approximately 25 to a few hundred. However, two studies had substantially larger sample sizes, specifically Graham and Gosling (2012, $n = 1{,}357$) and Youyou et al. (2015, $n = 17{,}662$). In the meta-analysis, studies with larger sample sizes are given more weight than the other studies in the meta-analysis, meaning they can have a greater influence on the overall findings. We therefore created a 90% winsorized data set, where sample sizes below and above the 5th and 95th percentiles were transformed to the value at the 5th and 95th percentile, respectively (Aguinis et al., 2013; Lipsey & Wilson, 2001). Here, sample sizes of 25 and 30 were winsorized to 40, while 1,357, and 17,662 were winsorized to 1,112.

Moreover, the study by Youyou et al. (2015) was also the only one within the set where the observers were friends with the targets, rather than zero-acquaintances. As such, observers would have had access to a much broader variety of cues to formulate their judgments. They would have observed their friend's Facebook profile over time and in a more ecologically valid setting, as opposed to a snapshot of a target's profile in a laboratory setting, and they would have had access to a variety of cues from their offline interactions with them. These differences in study design may also have a substantial impact on effect size estimates. Thus, we conducted a sensitivity analysis, where the Youyou et al. (2015) study was removed.

Overall, we reported three sets of meta-analytic results for each trait, one that included all studies, a second that comprised the winsorized data set, and a third reporting mean estimates with the Youyou et al.'s (2015) study removed. Table 1 displays the findings.

We took multiple steps to address publication bias. First, as outlined in our literature search, we conducted a broad search that encompassed published, unpublished, and grey literature. We also contacted all authors that had published reports appearing in our search, as well as others in the field for unpublished reports and data. We also examined funnel plot asymmetry and conducted a cumulative meta-analysis. These analyses were performed using the psychmeta package in R (Dahlke & Wiernik, 2019).

## *Study 2: Computer Prediction*

In total, there were 38 reports comprising 42 separate studies that met our inclusion criteria for the meta-analysis. These studies reported a total of 534 effect sizes. Similar to our approach for human perception, we extracted an effect size (i.e., Pearson's $r$ correlations) for each study for inclusion in the meta-analysis. However, in contrast to the human perception studies, the computer prediction studies report much more varied statistics and metrics. Therefore, in order to convert relevant statistics to correlations, we applied a number of transformations outlined as follows:

- Area under the receiver operating characteristic curve: Studies reporting area under curve were first converted to Cohen's $d$ using the method described by Ruscio (2008), and then converted to Pearson's $r$ using Rosenthal (1994). This was applicable for four studies ($k_s = 4$), with a total of 47 effect sizes ($k_{es} = 47$).

- Coefficient of determination (or $R^2$): Studies reporting results as $R^2$ were converted to correlations by taking the square root. This conversion was applied to four studies ($k_s = 4$) comprising a total of 55 effect sizes ($k_{es} = 55$).
- Diagnostic test indicators: Odds ratios were computed from studies that reported specificity (true negative rate) sensitivity (i.e., recall or true positive rate), precision (i.e., positive predicted values), and negative predicted values, or when sufficient information was provided to compute these statistics (Glas et al., 2003). Odds ratios were then converted to Cohen's *d* effect sizes (Borenstein et al., 2009), that were subsequently converted into correlations (Rosenthal, 1994). This process was applied to three studies ($k_s = 3$) comprising a total of 15 effect sizes ($k_{es} = 15$).
- Spearman's rho correlations were converted to Pearson's correlation in accordance with Gilpin (1993; $k_s = 1$ study, $k_{es} = 5$ effect sizes).

After performing the aforementioned transformations, we corrected artifactual variance across the effect sizes. Wiernik and Dahlke (2020) and Zhang (2022) found that meta-analyses rarely perform corrections (outside of industrial-organizational psychology), and those that do often report vague or limited detail. Moreover, artifacts can bias meta-analytic results and lead to inaccurate conclusions (Schmidt & Hunter, 2015; Wiernik & Dahlke, 2020). We therefore corrected all effect sizes for measurement error and sampling error artifacts. For measurement error, we used the test–retest reliabilities for the survey scales completed by the targets to calculate a corrected correlation. Since no reliabilities were available for computer predictions, we adopted a conservative approach and only corrected one variable (rather than both), in accordance with Harrer et al. (2021) and Schmidt and Hunter (2015)—a similar approach was adopted by Youyou et al. (2015) when comparing human perceptions and computer predictions. Sampling error was computed by calculating and correcting the standard error (*SE*) for each study following guidelines provided by Harrer et al. (2021). Effect sizes were also not transformed to Fisher's *z* scores in accordance with Schmidt and Hunter (2015) recommendations.

**Preliminary Analyses and Data Preparation.** The computer prediction reports varied substantially in their methods, experiments, and reporting of data. For instance, machine learning approaches often involve experimenting with different combinations of data and testing the performance of different types of algorithms in order to establish an optimal approach. Alternatively, some studies may test multiple approaches but only report the performance of a single algorithm and/or data type. Further, the accessibility of large data sets has led to repeated examination of publicly available data sets, where experiments are performed on a data set (or subset of that data) and are disseminated in different reports, or where workshops/competitions challenge researchers to compete for the best performing algorithm on a particular set of data (e.g., the myPersonality data set, Kosinski et al., 2013, and the PAN-AP-2015 data set, Rangel Pardo et al., 2015).

Traditional meta-analyses assume that effect sizes are independent, however the accumulation of these varied approaches created dependencies among the effect sizes. As such we conducted a multilevel meta-analysis—this is outlined in more detail in the Multilevel Meta-analysis section below. Nevertheless, the complex structure of our data set meant that there were numerous ways that data could be nested in the analysis. For example, where publicly available data sets create dependencies, prior meta-analyses have employed strategies to create single samples in the main analysis such as including the effect size derived from the largest sample (Chiang et al., 2022) or selecting the effect size derived from the best performing computer model (Marengo & Montag, 2020) prior to performing the multilevel analyses. For Chiang et al. (2022), such adjustments enabled their main analysis to focus on a three-level nested structure where individuals were nested within effect size, that were then nested within studies. For Marengo and Montag (2020), such adjustments enabled their main analysis to focus on a four-level nested structure that modeled effect sizes (Level 1), same-study effect sizes (Level 2), between-study variance (Level 3), and variance related to data sets (Level 4).

Consultation of these approaches in tandem with the meta-analytic literature indicated that there is no one-size-fits-all approach for performing multilevel meta-analyses and that existing guidance/simulation studies do not directly address the specifics of our data set. For example, Cheung (2014) argued that complex data sets may need four (or even more) levels, however (to our knowledge), research has not yet tested their effectiveness over three-level models. As such, we performed a preliminary four-level meta-analysis (with effect sizes at Level 1, within-study variance at Level 2, variance related to data sets at Level 3, and variance between studies at Level 4), however, the likelihood-ratio tests (LRTs) indicated that a four-level model was not better than a three-level model. Consequently, we made several adjustments to the data in order to perform a three-level meta-analysis. Following the guidelines provided by Senn (2009), we applied a rigorous set of criteria to identify overlapping samples from the meta-analysis. As such, samples were considered to be overlapping if they: (a) reported effect sizes that were based on overlapping sample subjects (i.e., subjects were part of the same overall data set), and (b) reported the same (or overlapping) digital traces derived from the same overall data set. For studies that aligned with these criteria, the study with the largest sample size was selected for inclusion. Then, studies that predicted personality using different types of data (again using the next largest samples available) were included, such that each study included predicted personality using different types of data. This resulted in the inclusion of three studies that used the myPersonality data set: Schwartz et al. (2013) predicted personality with language, Quercia et al. (2012) used number of friends, and Laleh and Shahram (2017) used likes. Consequently, 21 studies using the myPersonality data set were excluded as they included digital traces that overlapped with these studies. Further, two studies used the PsychFlickr dataset, both of which analyzed the same sample size (specifically, Cristani et al., 2013 and Guntuku et al., 2018). In this instance, Guntuku et al. (2018) was selected for inclusion on the basis that a wider variety of features were used to predict personality. Finally, two studies used a data set outlined in Golbeck et al. (2011), (specifically, Golbeck, 2016, and Golbeck et al., 2011). Here, Golbeck et al. (2011) was selected for inclusion based on the larger sample size and wider incorporation of features used to predict personality. These adjustments resulted in a total of 534 effect sizes ($k_{es} = 534$) extracted from 42 studies ($k_s = 42$) that were included in the multilevel meta-analysis.

**Multilevel Meta-Analysis.** As outlined above, extraction of effect sizes from the studies created many dependencies among the effect sizes. Dependencies can arise within studies when researchers test different algorithms and compare their performance or test the predictive validity of different features (or a combination of these two approaches). They can also arise between studies if effect sizes are derived from samples that share similarities in some way, for instance, the same cultural group or geographical region, or if the same research groups perform multiple studies. To handle such dependencies, we employed a multilevel modeling approach to perform the meta-analysis as it is purposely designed to model dependent effect sizes (in contrast to other meta-analytic approaches, that tend to assume independence among effect sizes, Cheung, 2014; Van den Noortgate et al., 2015). Multilevel approaches allow the inclusion of all available effect sizes, that maximize statistical power and preserve all information (Assink & Wibbelink, 2016). We also selected multilevel modeling over other approaches that handle dependencies among effect sizes such as robust variance estimation because we were interested in estimating the variance both within and between studies. Research comparing different methods for handling effect sizes has demonstrated that multilevel modeling is the best approach to use when this is the case—see Moeyaert et al. (2017). As such, the model comprised three levels, where Level 1 modeled the sampling variance of the effect sizes (i.e., the sampling attributable to individuals); Level 2 modeled the variance between effect sizes extracted from the same study (i.e., the variance within studies); and Level 3 modeled the variance between studies.

To perform the multilevel meta-analysis, we followed the guidelines and adapted the code provided by Assink and Wibbelink (2016). The multilevel meta-analysis was conducted in R (Version 4.2.1) with the *metafor* package (Viechtbauer, 2010), using a multilevel random effects model. The restricted maximum likelihood estimate was used to fit model parameters and the Knapp and Hartung (2003) adjustment was applied so that hypothesis tests were based on $t$ and $F$ distributions, rather than $Z$ distributions, thus improving Type I error (e.g., Assink & Wibbelink, 2016; Houben et al., 2015).

In order to ascertain the suitability of a three-level approach compared to a two-level analysis, two LRT were performed by freely estimating the variance at Level 3 and fixing the variance at Level 2 to zero (such that within-study variance is not modeled), and by freely estimating the variance at Level 2 and fixing the variance at Level 3 to zero (such that between-study variance is not modeled—see Assink & Wibbelink, 2016). The findings demonstrated that the fit of the three-level model was significantly better than the two-level model. Specifically, in the three-level model, the Akaike's information criteria (AIC) was −579.70, and the Bayesian information criteria (BIC) was −566.86. In contrast, the reduced two-level model where within-study variance = 0 showed an AIC of 183963.70, a BIC of 183972.26, and a LRT of 184545.40 ($p < .0001$). Additionally, the reduced two-level model where between-study variance = 0 had an AIC of 88.73, a BIC = 97.29, and an LRT = 670.43, ($p < .001$). These findings confirm that there is significant variability within-study variance (Level 2) as well as between-study variance (Level 3) and that the three-level model is statistically a better fit than the two-level model. In the Moderator Analyses section, we outline how moderator analyses were performed in attempt to explain this variance.

We assessed heterogeneity by examining the $Q$ statistic and the between- and within-study $I^2$, We also examined the between- and within-study estimates of $\tau^2$ because $I^2$, is influenced by the number of studies and is not an absolute measure of heterogeneity (Borenstein et al., 2017). All were calculated using the *metaphor* package in R (Viechtbauer, 2010) by following the guidelines and adapting the code by Assink and Wibbelink (2016).

**Publication Bias.** To assess publication bias, we inspected funnel plot asymmetry and conducted an Egger's regression test (Egger et al., 1997). It is worth noting that conventional tests for publication bias assume independence among effect sizes, meaning that such tests can inflate Type 1 error rates. Several simulation studies have evaluated various methods for examining publication bias in multilevel meta-analysis and provide approaches for addressing these issues (see Fernández-Castilla et al., 2021 and Rodgers & Pustejovsky, 2021). As such, we followed the guidelines of Fernández-Castilla et al. (2021) and adapted the Egger's regression to the structure of the multilevel meta-analysis. The funnel plot displays the effect sizes plotted against the inverse of standard errors (Egger et al., 1997; Sterne & Egger, 2001), where an asymmetrical plot indicates that publication bias might be present. The Egger's regression test provides an estimate of the asymmetry of the funnel plot. Again, all analyses were performed using the *metafor* package in R (Viechtbauer, 2010).

**Moderator Analyses.** In traditional meta-analyses that investigate the Big Five personality traits, researchers typically perform five separate meta-analyses (i.e., one for each trait). Conversely, the design and structure of the present multilevel analysis involves analyzing all five traits simultaneously, resulting in one overall effect size estimate for the effectiveness of digital traces at predicting personality. To examine the differences in convergent validity across all five traits, moderator analyses can be performed, where each trait is treated as a categorical moderator and effect sizes are grouped by each trait as a fixed effect in the multilevel model. Pairwise contrasts are then conducted to test for differences between across each of the traits.

We also extracted information from each study relating to moderators that may influence any heterogeneity among the effect sizes. These included a series of categorical moderators relating to data type, platform, type of publication, machine learning methods used, and a continuous moderator relating to publication year. These analyses were only conducted when at least five effect sizes were available for each category of the categorical moderators (publication year was the only continuous moderator examined that applied to all studies/effect sizes).

Data type moderators were coded into five categories including language, picture, metadata, likes, and multiple data. Both the language and picture categories included varied types of data that comprised language or pictures in some form or another. For example, pictures can be analyzed in various ways by extracting different features and training a computer model using them. Those features may include colors, content of the pictures, depth of field, etc. Because machine learning enables analysis of such fine-grained detail and features, it would not be possible to include each of these features separately in a meta-analysis, as it is rare that two studies would use the same combination of features so that they could be compared. Thus, these broad categories were used to encompass features that related to such overarching categories. The metadata category encompassed studies that used any form of metadata to predict personality, for instance, number of likes, number of updates, number of friends, etc. The likes category included studies that analyzed the different types of likes that could be used to predict personality, and the multiple data type included studies that used combinations of data that could otherwise be classified under multiple categories, for instance, using likes, status updates, and number of friends to predict personality.

Platform moderators represented the different types of platforms (i.e., different types of social media, devices, or applications) that digital traces were extracted from to predict personality. They were coded into twelve different categories that included Spotify, X, Facebook, Flickr, Sina Weibo, XING, smartphones, a money management application, personal computers, videos, online service logs, and multiple platforms (i.e., where data traces extracted from more than one platform were used to predict personality).

We also coded studies in accordance with their "publication type," that is, whether the study was published in a journal or as part of conference proceedings.[5] Most of the journal-based studies in our set were published in psychology journals, and most of the conference-based studies were published in computer science conferences. Both fields are highly disparate and have different criteria for what constitutes publishable research, for instance, psychology is often concerned with explaining psychological phenomena, whereas computer science is typically concerned with data-driven predictions (Hinds & Joinson, 2019; Yarkoni & Westfall, 2017). Examining the different types of publication will therefore show whether publication type has a moderating effect on personality prediction.

Moderators for the type of machine learning method used were coded into three categories that covered the type of approach that researchers used to validate the results of the computer models. The different approaches include the holdout method (where the data are randomly split into a larger training set and a smaller testing set, upon which the models are trained and tested), *k*-fold validation (where data are split into training/test sets but where the process is repeated *k* times, analyses are performed on each fold and then combined to form an overall estimate), and no cross validation (where no steps were taken to validate the data—see Marengo & Montag, 2020; Marengo & Settanni, 2019, for related work). Finally, the year of publication was included as a continuous moderator that was centered around the year of the earliest report (published in 2011).

Moderator analyses were performed by conducting a series of meta-regression analyses for multilevel modeling. To do this, we first tested the potential moderating effect of multiple variables separately in univariate models (i.e., univariate models for data type, platform, type of publication, machine learning methods, and publication year). However, given that variables can be interrelated (that can lead to substantial multicollinearity, Hox et al., 2017), we then tested multiple moderators in a single, multivariate model that analyzed the significant moderators identified in the univariate analyses simultaneously (following the guidelines by Assink & Wibbelink, 2016; Hox et al., 2017). Significant omnibus *F* tests of coefficients were used to determine differences among coefficients in the model. These follow an *F*-distribution where degrees of freedom are based on the number of coefficients tested ($df_1$) and the number of effect sizes minus the total number of coefficients in the model including the intercept ($df_2$), thus significant individual regression coefficients indicate moderating effects (Assink & Wibbelink, 2016; Viechtbauer, 2010). Finally, we calculated 95% prediction intervals for each moderator group. These provide a range from which future true-effect sizes from randomly selected populations can be expected to fall (IntHout et al., 2016). In contrast to 95% confidence intervals (that provide insight into the precision of the mean effect size estimation), prediction intervals provide information on the heterogeneity in the overall distribution of true-effect sizes.

## Transparency and Openness

We adhered to the Preferred Reporting of Items for Systematic Reviews and Meta-Analysis 2020 guidelines for systematic reviews (Page et al., 2021). All meta-analytic data and analysis code are available at https://osf.io/9cndh/. Data were analyzed using the R packages *psychmeta* (Dahlke & Wiernik, 2019) and *metafor* (Viechtbauer, 2010), Version 4.2.1. The review was not preregistered.

# Results

Over the following sections, we organize our findings in line with our main research questions. That is, we first analyze the main meta-analytic findings in terms of *convergent validity* (Research Question 1) and explore differences in the *identifiability* of the Big Five personality traits (Research Question 2). Then, we report our moderator analyses following Schmidt and Hunter's (2015) procedure for the human perception studies, and following guidelines provided by Assink and Wibbelink (2016) for the computer prediction studies. These analyses enabled us to disentangle whether certain types of digital data are more indicative of personality than others (i.e., *predictability,* Research Question 3). Moreover, in synthesizing the diverse range of evidence that falls outside of effect sizes that can be included in meta-analyses, we summarize study characteristics, data, and findings in the Supplemental Materials, and Table 1 presents a taxonomy of the diverse range of platforms and types of digital data that have been studied to date.

## Study 1: Human Perception

### *Primary Analyses*

First, we performed meta-analyses on each personality trait. These results are presented in Table 2. Overall, the mean observed effect sizes were low-to-moderate, ranging from $\bar{r} = 0.29$ (for Conscientiousness) to $\bar{r} = 0.42$ (for Openness). Yet, when corrected for measurement error, they increased to $\rho = 0.38$ (for Neuroticism) to $\rho = 0.57$ (for Openness). The confidence intervals were overlapping for Openness and Extraversion (the traits with the highest self-other agreement) and for Conscientiousness, Agreeableness, and Neuroticism (the traits with the lowest self-other agreement) and none included zero. Our analyses also indicated substantial heterogeneity across all traits, as evidenced by multiple measures displayed in Table 2. Although the credibility intervals did not include zero (with the exception of the outlier analysis for Agreeableness), they were generally quite broad. There was some variability around the SDρ values across all traits. The I2 values (that indicate the percentage of variability in effect sizes due to heterogeneity rather than sampling error; Higgins & Thompson, 2002) were high across all traits (ranging from 75.42 for

[5] Although our search included grey literature, including unpublished reports and dissertations, no such reports were included in these analyses as they contained nonindependent data. Thus, the "publication type" moderator comprised only journal and conference proceedings categories.

**Table 2**
*Meta-Analyses for Human (Self-Other) Perception Across All Studies*

| Trait and moderator | $k_s$ | $n$ | $\bar{r}$ | $SD\bar{r}$ | ρ | $SD$ρ | [95% CI] | [80% CR] | %Var | $Q$ ($df$) | $I^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Openness | **29** | **24,052** | **0.42** | **0.17** | **0.57** | **0.10** | **[0.53, 0.61]** | **[0.44, 0.70]** | **11.75** | **238.26 (28)*****| **88.25** |
| Openness[a] | 29 | 7,282 | 0.24 | 0.18 | **0.46** | 0.21 | [0.37, 0.55] | [0.19, 0.74] | 19.74 | 141.89 (28)*** | 80.26 |
| Openness[b] | 28 | 6,390 | 0.18 | 0.15 | **0.39** | 0.23 | [0.29, 0.49] | [0.09, 0.69] | 24.07 | 112.17 (27)*** | 75.93 |
| Social media (Level 1)[a] | 19 | 4,103 | 0.33 | 0.17 | **0.48** | 0.21 | [0.37, 0.59] | [0.21, 0.76] | 15.07 | 119.48 (18)*** | 84.93 |
| Social media (Level 1)[b] | 18 | 2,966 | 0.26 | 0.14 | **0.41** | 0.24 | [0.27, 0.54] | [0.08, 0.73] | 19.05 | 89.23 (17)*** | 80.95 |
| Nonsocial media (Level 1)[a] | 10 | 3,179 | 0.12 | 0.25 | **0.37** | 0.21 | [0.19, 0.55] | [0.08, 0.65] | 33.38 | 26.96 (9)** | 66.62 |
| Facebook (Level 2)[a] | 10 | 2,868 | 0.35 | 0.16 | **0.50** | 0.15 | [0.3, 0.62] | [0.29, 0.70] | 18.41 | 48.90 (9)*** | 81.59 |
| Facebook (Level 2)[b] | 9 | 1,731 | 0.25 | 0.11 | **0.39** | 0.15 | [0.24, 0.53] | [0.18, 0.60] | 33.89 | 23.60 (8)* | 66.11 |
| Other platforms (Level 2)[a] | 9 | 1,235 | 0.28 | 0.19 | **0.44** | 0.36 | [0.14, 0.73] | [−0.07, 0.94] | 12.09 | 66.20 (8)*** | 87.92 |
| Conscientiousness | **29** | **24,040** | **0.29** | **0.12** | **0.39** | **0.07** | **[0.36, 0.42]** | **[0.30, 0.48]** | **24.58** | **113.92 (28)***** | **75.42** |
| Conscientiousness[a] | 29 | 7,270 | 0.17 | 0.13 | **0.31** | 0.13 | [0.24, 0.37] | [0.14, 0.47] | 38.14 | 73.42 (28)*** | 61.86 |
| Conscientiousness[b] | 28 | 6,378 | 0.13 | 0.12 | **0.26** | 0.13 | [0.19, 0.33] | [0.08, 0.43] | 44.80 | 60.28 (27)*** | 55.21 |
| Social media (Level 1)[a] | 19 | 4,143 | 0.25 | 0.11 | **0.35** | 0.11 | [0.28, 0.42] | [0.20, 0.49] | 40.03 | 44.96 (18)*** | 59.97 |
| Social media (Level 1)[b] | 18 | 3,016 | 0.21 | 0.10 | **0.31** | 0.12 | [0.23, 0.39] | [0.15, 0.47] | 45.50 | 37.36 (17)* | 54.50 |
| Nonsocial media (Level 1)[a] | 10 | 3,127 | 0.06 | 0.07 | **0.15** | 0.08 | [0.04, 0.26] | [0.04, 0.26] | 73.00 | 12.33 (9) | 27.00 |
| Facebook (Level 2)[a] | 10 | 2,908 | 0.29 | 0.09 | **0.38** | 0.08 | [0.30, 0.46] | [0.27, 0.49] | 43.75 | 20.57 (9)* | 56.25 |
| Facebook (Level 2)[b] | 9 | 1,781 | 0.25 | 0.09 | **0.35** | 0.11 | [0.24, 0.46] | [0.20, 0.50] | 45.71 | 17.50 (8)* | 54.29 |
| Other platforms (Level 2)[a] | 9 | 1,235 | 0.15 | 0.11 | **0.24** | 0.13 | [0.09, 0.38] | [0.05, 0.42] | 51.33 | 15.59 (8) | 48.67 |
| Extraversion | **30** | **24,124** | **0.40** | **0.14** | **0.52** | **0.08** | **[0.48, 0.55]** | **[0.41, 0.63]** | **16.01** | **181.15 (29)***** | **83.99** |
| Extraversion[a] | 30 | 7,295 | 0.26 | 0.17 | **0.44** | 0.17 | [0.37, 0.51] | [0.22, 0.65] | 23.36 | 124.14 (29)*** | 76.64 |
| Extraversion[b] | 29 | 6,462 | 0.22 | 0.16 | **0.40** | 0.17 | [0.33, 0.48] | [0.18, 0.62] | 28.13 | 99.54 (28)*** | 71.87 |
| Social media (Level 1)[a] | 19 | 4,093 | 0.33 | 0.16 | **0.48** | 0.21 | [0.38, 0.59] | [0.21, 0.76] | 15.20 | 118.42 (18)*** | 84.80 |
| Social media (Level 1)[b] | 18 | 2,966 | 0.26 | 0.14 | **0.41** | 0.24 | [0.27, 0.54] | [0.08, 0.73] | 19.05 | 89.23 (17)*** | 80.95 |
| Nonsocial media (Level 1) | 11 | 3,211 | 0.15 | 0.15 | **0.36** | 0.16 | [0.23, 0.49] | [0.15, 0.57] | 34.30 | 29.16 (10)** | 65.70 |
| Facebook (Level 2)[a] | 10 | 2,908 | 0.40 | 0.12 | **0.50** | 0.13 | [0.40, 0.61] | [0.32, 0.69] | 18.29 | 49.20 (9)*** | 81.71 |
| Facebook (Level 2)[b] | 9 | 1,781 | 0.36 | 0.11 | **0.48** | 0.15 | [0.35, 0.62] | [0.27, 0.70] | 23.75 | 33.69 (8)*** | 76.25 |
| Other platforms (Level 2)[a] | 9 | 1,235 | 0.24 | 0.12 | **0.33** | 0.20 | [0.15, 0.51] | [0.06, 0.61] | 27.46 | 29.14 (8)*** | 74.55 |
| Agreeableness | **29** | **24,052** | **0.34** | **0.13** | **0.46** | **0.11** | **[0.42, 0.51]** | **[0.32, 0.61]** | **11.83** | **236.75 (28)***** | **88.17** |
| Agreeableness[a] | 29 | 7,282 | 0.21 | 0.17 | **0.37** | 0.23 | [0.27, 0.46] | [0.07, 0.66] | 16.61 | 168.54 (28)*** | 83.39 |
| Agreeableness[b] | 28 | 6,390 | 0.17 | 0.16 | **0.32** | 0.25 | [0.21, 0.42] | [−0.01, 0.64] | 18.31 | 147.47 (27)*** | 81.70 |
| Social media (Level 1)[a] | 19 | 4,143 | 0.28 | 0.13 | **0.40** | 0.16 | [0.31, 0.49] | [0.18, 0.62] | 22.77 | 79.05 (18)*** | 77.23 |
| Social media (Level 1)[b] | 18 | 3,106 | 0.24 | 0.12 | **0.35** | 0.18 | [0.25, 0.46] | [0.12, 0.59] | 28.41 | 59.84 (17)*** | 71.59 |
| Nonsocial media (Level 1) | 10 | 3,139 | 0.11 | 0.17 | **0.25** | 0.36 | [−0.02, 0.53] | [−0.25, 0.75] | 11.30 | 79.63 (9)*** | 88.70 |
| Facebook (Level 2)[a] | 10 | 2,908 | 0.32 | 0.12 | **0.43** | 0.16 | [0.30, 0.55] | [0.21, 0.64] | 16.69 | 53.91 (9)*** | 83.31 |
| Facebook (Level 2)[b] | 9 | 1,781 | 0.28 | 0.12 | **0.38** | 0.18 | [0.22, 0.54] | [0.13, 0.64] | 20.75 | 38.55 (8)*** | 79.25 |
| Other platforms (Level 2)[a] | 9 | 1,35 | 0.18 | 0.11 | **0.29** | 0.17 | [0.12, 0.47] | [0.05, 0.53] | 41.08 | 19.47 (8)* | 58.92 |
| Neuroticism | **28** | **23,968** | **0.29** | **0.12** | **0.38** | **0.07** | **[0.35, 0.41]** | **[0.28, 0.48]** | **21.25** | **127.05 (27)***** | **78.75** |
| Neuroticism[a] | 28 | 7,256 | 0.17 | 0.15 | **0.32** | 0.17 | [0.25, 0.40] | [0.10, 0.55] | 26.90 | 100.36 (27)*** | 73.10 |
| Neuroticism[b] | 27 | 6,306 | 0.13 | 0.14 | **0.28** | 0.21 | [0.19, 0.38] | [0.01, 0.55] | 28.16 | 92.32 (26)*** | 71.84 |
| Social media (Level 1)[a] | 19 | 4,143 | 0.23 | 0.14 | **0.33** | 0.14 | [0.25, 0.42] | [0.15, 0.52] | 29.55 | 60.91 (18)*** | 70.45 |
| Social media (Level 1)[b] | 18 | 3,106 | 0.18 | 0.14 | **0.29** | 0.17 | [0.19, 0.40] | [0.06, 0.52] | 32.02 | 53.10 (17)* | 67.99 |
| Nonsocial media (Level 1)[a] | 9 | 3,055 | 0.08 | 0.13 | **0.26** | 0.30 | [0.00, 0.52] | [−0.16, 0.69] | 20.66 | 38.73 (8)* | 79.35 |
| Facebook (Level 2)[a] | 10 | 2,908 | 0.28 | 0.11 | **0.37** | 0.11 | [0.28, 0.47] | [0.23, 0.52] | 31.39 | 28.67 (9)** | 68.61 |
| Facebook (Level 2)[b] | 9 | 1,781 | 0.24 | 0.12 | **0.36** | 0.16 | [0.21, 0.50] | [0.14, 0.57] | 30.29 | 26.41 (8)** | 69.71 |
| Other platforms (Level 2)[a] | 9 | 1,235 | 0.10 | 0.13 | **0.18** | 0.16 | [0.01, 0.34] | [−0.05, 0.40] | 45.36 | 17.64 (8)*** | 54.64 |

*Note.* Bold values indicate the main findings for each trait and corrected point estimate for each analysis. $k_s$ = number of independent samples; $n$ = number of targets in a sample; $\bar{r}$ = mean observed correlation; $SD\bar{r}$ = standard deviation of the observed correlation; ρ = corrected point estimate; $SD$ρ = standard deviation of corrected correlation; 95% CI = confidence intervals around ρ; 80% CR = credibility intervals around ρ; %Var = percentage of observed variance accounted for by statistical artifacts; $Q$ ($df$) = heterogeneity $Q$ statistic; $I^2$ = $I^2$ heterogeneity.
[a] Meta-analytic results for winsorized analyses. [b] Meta-analytic results for the sensitivity analyses with the Youyou et al. (2015) study removed.
* $p < .05$. ** $p = .001$. *** $p < .001$.

Conscientiousness to 88.25 for Openness), and the Q statistic was significant across all traits (p < .001). The percentage of variance explained by artifacts was generally low across all traits (from Openness = 11.75% to Conscientiousness = 24.58%)—Schmidt and Hunter (2015) indicated that variance less than 75% suggests the presence of moderators.

### Sensitivity Analyses and Publication Bias

To investigate whether estimates were influenced by outliers, we performed two sets of sensitivity analyses, one where the dataset was winsorized, and another where the Youyou et al.'s (2015) study was removed (to address any influence exerted on the mean effect size due to the disproportionately large sample size, as well as differences in study design relating to observers being friends with targets and likely knowing them both on- and offline). In the first analysis, we winsorized 90% of the data set where samples of 1,357 (Graham & Gosling, 2012) and 17,662 (Youyou et al., 2015) were winsorized to 1,112 and samples of 25 (Darbyshire et al., 2016), and 30 were winsorized to 40 (Waggoner et al., 2009)—the results are displayed in Table 2. The meta-analyses were repeated and demonstrated lower mean effect sizes. These ranged from $\bar{r} = 0.17$ (for Conscientiousness and Neuroticism) to $\bar{r}$ = 0.26 (for Extraversion), and ρ = 0.31 (for Conscientiousness) to ρ = 0.46 (for Openness) when corrected for measurement error. The confidence intervals across all five traits overlapped, ranging from CI [0.24, 0.37] for Conscientiousness to CI [0.37, 0.55] for Openness and none included zero. Moreover, the outlier analysis demonstrated further lowered effect sizes, ranging from $\bar{r}$ = 0.13 (for Conscientiousness and Neuroticism) to $\bar{r}$ = 0.22 (for Extraversion), and ρ = 0.28 (for Conscientiousness) to ρ = 0.40 (for Extraversion) when corrected for measurement error. The confidence intervals across all five traits overlapped, ranging from CI [0.19, 0.33] for Conscientiousness, to CI [0.23, 0.49] for Openness, and none included zero.

Across both analyses, the credibility intervals were broader than the primary meta-analyses and did not include zero (apart from the outlier analysis for Agreeableness). Further, the outlier analyses resulted in broader credibility intervals than the winsorized analyses for all traits apart from Openness, demonstrating that the parameter values are more widely distributed when the mean rho is not influenced by Youyou et al.'s (2015) study. Further, across both analyses, the *SD*ρ. for all traits demonstrated higher variability than the primary analyses, $I^2$ were lower (but still indicated substantial heterogeneity), and all $Q$ statistics were significant ($p$ < .001). Finally, the percentage of variance explained by artifacts was high across all traits, but still indicated the presence of moderators.

Together, the primary and sensitivity analyses indicate that Youyou et al.'s (2015) study inflated the mean effect sizes across all traits (their reported effect sizes were moderate, ranging from ρ = 0.39 to ρ = 0.59). This is also likely due to their use of friends as observers and the access to both online and offline cues when rating targets' personalities. Moreover, the effect sizes reported by Youyou et al. (2015) align with research demonstrating that acquainted observers demonstrate higher convergent validity than zero-acquaintance observers (e.g., Connelly & Ones, 2010). Indeed, the range of effect sizes reported by Youyou et al. (2015) is similar to friends' ratings of targets reported by Connelly and Ones (2010). For completeness, we repeated these tests in our moderator analyses; these are also displayed in Table 2.

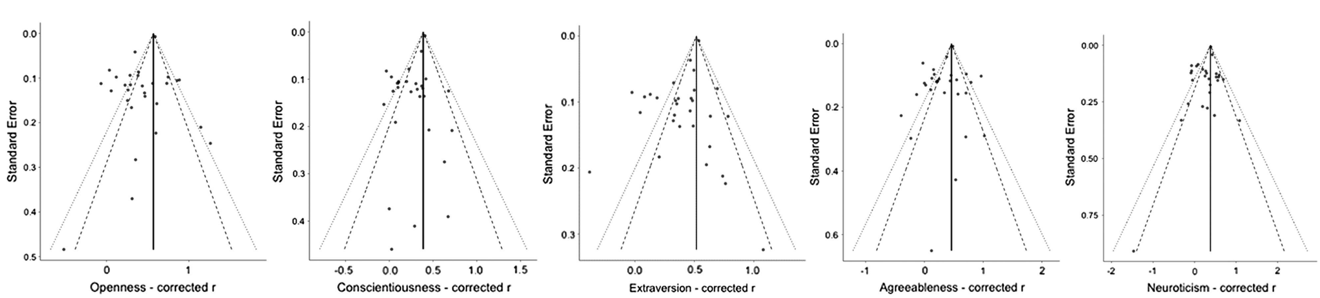
These findings are broadly consistent with other meta-analyses comparing zero-acquaintance-target ratings. These studies found lower self-other agreement for Neuroticism, that is low in visibility, compared to other traits (Connelly & Ones, 2010; Oh et al., 2011) and higher self-other agreement for Extraversion, that is high in visibility (Connelly & Ones, 2010). Our findings for the other traits were somewhat mixed in comparison to previous work, for instance, Oh et al. (2011) found that self-other agreement was highest for Conscientiousness, and Connelly and Ones (2010) found that observers could not perceive Openness as effectively relative to other traits. That said, the present research is markedly distinct from the previous work in that the observers were exposed to and utilized vastly different types of cues. For instance, although Connelly and Ones's (2010) study did include some digital traces in their effect sizes, most studies were conducted in nondigital contexts, such as job interviews and group interactions.

Next, we performed tests for publication bias. Inspection of the funnel plots revealed some asymmetry across all traits indicating the presence of publication bias (see Figure 2). For Openness, Conscientiousness, Agreeableness, and Neuroticism, there were a small number of studies to the left of the mean, whereas for Extraversion there was a study to the right. The findings for the former traits run counter to expectations for publication bias tests, where publication bias would typically be evidenced by a positive association between standard error and effect size (which would be displayed to the right of the mean, rather than the left). It is worth noting that most studies were clustered toward the top of the funnel, with only a few toward the base. The cumulative meta-analyses demonstrated a negative shift in effect sizes with the addition of smaller studies, that would also indicate the presence of publication bias for each trait. However, given the disproportionately large effect/sample size, and methodology applied in Youyou et al.'s (2015) work (that may influence these results), we ran a second set of cumulative meta-analyses with this study removed. These findings were mixed; Openness, Conscientiousness, and Agreeableness displayed some drift but did reach stability, ρ = 0.40, CI [0.19, 0.62] (positive drift); ρ = 0.26, CI [0.19, 0.33] (negative drift); and ρ = 0.32, CI [0.21, 043] (negative drift), respectively. In contrast, the effect sizes for Extraversion and Neuroticism were generally stable. Based on these findings, the extent that publication bias may be present is unclear. As such, we interpret these findings with caution, recognizing the small $k_s$ overall, and the large presence of heterogeneity. It is therefore likely that more studies would need to be conducted to unpack these effects further.

### Hierarchical Subgroup Analyses

We explored potential moderator effects by applying Schmidt and Hunter's (2015) hierarchical subgrouping approach that involves dividing studies into subgroups, meta-analyzing the subgroups, and dividing the subgroups into smaller groups recursively until

**Figure 2**
*Funnel Plots for Human (Self-Other) Perception Across Each Personality Trait*



acceptable homogeneity of effect sizes is achieved. Subgrouping is an ideal approach in instances like this when moderators are categorical or dichotomous. Alternative approaches such as meta-regression can capitalize on chance, creating inflated Rs (Raudenbush et al., 2006) and typically rely on low statistical power, meaning that the weights for most real moderators will be nonsignificant (Schmidt & Hunter, 2015). That said, hierarchical subgroup analysis substantially reduces the number of studies included in each step and increases the sampling error within each group. Second-order sampling error is therefore likely to be increased at each level of the analysis.

We were limited in the analyses we could perform because of the relatively small number of studies included in the overall set, and due to the diversity of platforms and types of digital data investigated within those studies. For instance, our data set comprised eleven different types of platforms (across thirty studies), and in several cases, platforms were only reported in one study. Moderator analyses based on a small $k_s$ is a common problem in meta-analytic investigations (Schmidt & Hunter, 2015), and this is pronounced in this instance as researchers have been investigating diverse ways that personality can be assessed online. Consequently, our subgrouping approach was relatively broad and included studies that are likely to be heterogeneous (e.g., different social media platforms are likely to vary but were pooled as a "social media" subgroup).

To that end, we examined moderator effects across two hierarchical levels. First, studies were partitioned into two "data source" subgroups—social media and nonsocial media (Level 1). Then, social media was partitioned into two further subgroups—"Facebook" and "Non-Facebook" (Level 2). The results for each of the subgroup analyses are displayed in Table 2 underneath the primary and sensitivity analyses. Despite the limitations imposed by the number of subgroups examined, analyzing these groups allows us to attain some understanding of how different types of data (across social media or other platforms) may help or hinder self-other agreement. As such, our analyses should be treated as exploratory, and our findings should be interpreted with caution. In our broader, narrative discussion of current work, we reflect on how existing methods, and the reasons why different platforms/types of data may contribute toward existing heterogeneity. We then use this discussion to inform future avenues of research in the General Discussion.

**Level 1 (Data Source) and Level 2 (Social Media) Analyses.** Overall, both subgroup analyses demonstrated higher self-other agreement for judgments made from social media accounts than other types of media (Level 1), and from Facebook than other platforms across all traits[6] (Level 2). Across both sets of subgroup analyses and for all traits, confidence intervals overlapped, with the exception of Conscientiousness in Level 1 (winsorized data). Further, the confidence intervals generally did not cross zero (with the exceptions of nonsocial media in Agreeableness and Neuroticism). The credibility intervals also remained broad across all subgroup analyses and generally did not include zero (with the exceptions of nonsocial media in Agreeableness and Neuroticism, Level 1).

The *SD* values were mixed across both levels. In the Level 1 data source analyses, *SD*ρ values for social media and nonsocial media in Conscientiousness displayed some reduction. Alternatively, *SD*ρ values for Openness demonstrated no change, and the values for Extraversion, Agreeableness, and Neuroticism showed some increases and decreases in variability. In the Level 2 social media analyses, the *SD*ρ values for Facebook and other platforms were further lowered for Neuroticism and Extraversion. The other *SD*ρ values were mixed. Further reductions were evident for Facebook in terms of Openness and Conscientiousness, as well as for nonsocial media and Agreeableness. In contrast, *SD*ρ values demonstrated no change for Facebook and Agreeableness and were higher for other platforms in both Openness and Conscientiousness. Moreover, with the exception of the subgroup analysis for nonsocial media at Level 1 for Conscientiousness, all $I^2$ values indicated moderate to substantial heterogeneity. The percentage of variance explained by artifacts increased but remained low across all traits. Additionally, the $Q$ statistic was significant across all analyses ($p < .05$).

As outlined above, the presence of heterogeneity within these analyses is not surprising, given the diversity of data types explored in the studies. It is also important to note that these results may be subject to second-order sampling error, that limits the insights that can be derived from the analyses. More studies on self-other agreement across different platforms is therefore needed to further unpack the extent of moderator effects human personality perception using digital data.

### Discussion

Overall, our meta-analysis highlights that convergent validity was modest for all five traits. This evidence offers support that digital traces reflect people's personalities, and that strangers can successfully infer others' personalities based on these cues. These findings are broadly consistent with the previous personality research where self-other agreement tends to be lower for Neuroticism and higher for Extraversion (e.g., Connelly & Ones, 2010; Tskhay & Rule, 2014). Extraversion is high in visibility and low in evaluativeness, and therefore is likely to be readily perceivable in targets' interactions with friends and communications relating to social activity/socialization (e.g., Bowden-Green et al., 2020). In contrast, Neuroticism is low in visibility, and our findings support the idea that it is more difficult to perceive than other traits. Research has demonstrated that individuals high in Neuroticism weigh up the risks of disclosing information online because they may expose information that could be used against them (Loiacono, 2015). Moreover, they are often self-conscious of how they appear and therefore may present an "ideal" or "false" self-online (e.g., Bowden-Green et al., 2021; Michikyan et al., 2015).

Agreeableness and Conscientiousness, that are both moderate in observability and evaluativeness also had lower self-other agreement. Similar to individuals high in Neuroticism, conscientious individuals are also cautious in their self-presentation online (Seidman, 2013), they tend to spend less time using social media, they infrequently post photos (Amichai-Hamburger & Vinitzky, 2010; Ryan & Xenos, 2011) and comment on others' posts (Lee et al., 2014). Agreeable individuals are often concerned with using social media to seek acceptance and maintain connection, presenting their actual self rather than seeking attention (Seidman, 2013). Such behaviors across individuals high in Neuroticism, Conscientiousness, and Agreeableness may therefore hinder observers' abilities to perceive

[6] The only exception to this was the subgroup analysis for Openness where the Youyou et al. (2015) study was removed from the Facebook category (Level 2) where ρ was lower than other platforms (ρ = 0.25, and ρ = 0.28, respectively).

these traits effectively (particularly when observers are strangers), as they are unlikely to be readily observable. These findings support theoretical accounts that suggest that internalized traits are more difficult to perceive than externalized traits (Funder, 1995; John & Robins, 1993) and extend patterns of trait visibility to online environments.

Alternatively, Openness (which is typically low in visibility and high in evaluativeness) appears to be more observable from digital traces. Previous research has argued that individuals high in Openness refrain from revealing religious or political beliefs, creative ideas, or emotions to strangers (e.g., Connelly & Ones, 2010)—this makes sense in face-to-face interactions as people try to adhere to social norms and avoid controversy. Conversely, in online environments people often reveal such information, through their profile data, interactions, etc. CMC is often found to make such disclosures easier because individuals are physically separated from their audience (e.g., Chan-Olmsted et al., 2013; Kiesler et al., 1984; Peter & Valkenburg, 2006). Likewise, people are generally communicating with people they know in a domain where strangers can witness these interactions (that would not typically occur in offline interactions to the same degree). These findings therefore suggest that aspects such as visibility and evaluativeness may influence how effectively traits can be *identified* (i.e., whether they are "good traits"; Funder, 1995, 2012) and that CMC may contribute toward improving the visibility of Openness.

Although our subgroup analyses were too limited to draw conclusions about the moderating effects of different platforms and types of data, we tentatively speculate how these aspects may influence convergent validity, the identifiability and predictability of different personality traits in the hope that future research could unpack these uncertainties. As such, based on our findings, it seems likely that judgements made from social media may be more effective from those made from other types of media/platforms. Social media can provide access to "good information" (Funder, 1995) in that there are more cues generated by targets that are perceivable to observers; targets can publish different types of content (e.g., photos, videos, likes, links), and observers can utilize these cues, as well as draw upon content from the targets' network (e.g., comments, photos). Such possibilities are typically not available in other forms of media (or at least not to the same extent), reducing the amount of information an observer has access to.

Explorations of other types of media/platforms were limited in that researchers often focused on one type of cue (e.g., language in the form of email addresses, emails, chats). Although this enables researchers to hone in on the *predictability* of specific digital traces, observers have less information to formulate judgments. Similarly, targets have different means and motives for using social media compared to other media. For instance, emails are private, and the topic being discussed may not reveal anything personal or directly related to the target. Alternatively, although targets may use social media for a variety of purposes, socialization and exchanging information amongst networks of friends are often central to their activities (Quan-Haase & Young, 2010; Whiting & Williams, 2013). Likewise, observers can access information about the target produced by other people (i.e., their friends, family, colleagues), that may increase the validity of the cues because targets are unable to manipulate or control how they are perceived (known as *warranting theory,* Walther & Parks, 2002). Thus, it seems reasonable to infer that having access to a wider range of cues will contribute toward higher self-other agreement. Other types of media such as personal websites, blogs, or vlogs may offer similar advantages, although from existing work we do not know how much detail was published (or what types of cues were used by observers, etc.) on these platforms as such information was not studied or reported.

It also seems likely that convergent validity will vary across social media platforms, given that individuals often use multiple platforms simultaneously (Brandtzæg, 2012) in order to meet different needs (Pelletier et al., 2020; Phua et al., 2017) and change the way they present themselves across various platforms (Davidson & Joinson, 2021). For example, individuals often use Facebook to communicate, socialize with and seek information about others within their network whom they know offline (e.g., N. B. Ellison et al., 2007; Krause et al., 2014; Orchard et al., 2014), whereas on X, interactions can occur with close ties (such as family and friends) as well as weak ties with individuals whom they have never met or interacted with before (such as celebrities, politicians, and members of the public, Jin & Phua, 2014; Phua et al., 2017; Quan-Haase et al., 2015). Alternatively, on LinkedIn individuals seek to establish professional connections (e.g., Buettner, 2017; Carmack & Heiss, 2018), fulfilling both personal and social needs (i.e., maintaining and strengthening connections, by learning about people they have met, or plan to meet, Florenthal, 2015; Smith & Watkins, 2020). Thus, convergent validity may be higher on Facebook than X because individuals share more (and personal) information amongst their private networks. Similarly, it may be easier to *identify* certain traits on one platform than another. For instance, Conscientiousness may be much more visible on LinkedIn than on X as individuals work hard to develop/maintain professional relationships and display expertise about their field of work. Conscientiousness may therefore be displayed via informative posts and comments on others' posts. Such cues may not be as indicative of Conscientiousness on X if for instance, individuals are using the platform to keep updated with news. Future work could therefore seek to investigate this.

The identifiability of different traits may also vary across media, and this may be compounded by the affordances of different platforms. For instance, individuals high in Neuroticism may refrain from conveying their anxieties on social media but may do so freely within an online community. Online communities are spaces where individuals seek support and connection with others, and the opportunity to interact anonymously can increase individuals' self-disclosure (e.g., Li et al., 2015; Wright, 2016). It also seems likely that affordances such as interaction, or whether communications are public or private could influence self-other agreement. In reality, strangers will often formulate judgments about others whom they have never met before via direct communications (e.g., on social media, online dating). The opportunity to communicate directly may help observers to develop a clearer impression of the person by asking direct questions about who they are and their life. Likewise, the ability to interact publicly versus privately may also influence self-other agreement for different traits. For instance, in private one-to-one messaging, introverts may be more communicative, and individuals high in Neuroticism may feel more secure to disclose personal information. Finally, the usability and design affordances of different sites could also influence the predictability of certain cues. On social media, posts that are highly "liked" may influence impressions more than posts that are not liked. In online communities, reputation systems (where individuals acquire certain levels of status or scores that indicate their trustworthiness from their online behaviors, e.g., Hendrikx et al., 2015; Jensen et al., 2002; Rough et al., 2021) could be highly informative for strangers that have never interacted with a target before. Future research could therefore explore these possibilities in further detail.

## Study 2: Computer Prediction

### *Primary Analysis*

The estimated overall meta-analytic effect of predicting personality using digital traces was significant, $\rho = 0.30$ ($SE = 0.03$; 95% CI [0.24, 0.36]). Table 3 displays the results for all analyses. Figure 3 displays a caterpillar plot of the included effect sizes, that range from $\rho = -0.39$ to $\rho = 1.09$. The result of the $Q$ test indicated significant heterogeneity among the effect sizes, $Q(533) = 454858.11$, $p < .001$. There was significant heterogeneity both within studies (Level 2), $\tau_1^2 = 0.01$, 95% CI [0.01, 0.01], and between studies (Level 3), $\tau_2^2 = 0.04$, 95% CI [0.03, 0.07]. Further, the total variance distributed across the three levels ($I^2$) indicated that 0.09% of the variance was attributed to the sampling variance at Level 1, 22.21% of the variance was attributed to the within-study variance at Level 2, and 77.70% of the variance was attributed to between-study variance at Level 3.

### *The Identifiability of Different Personality Traits*

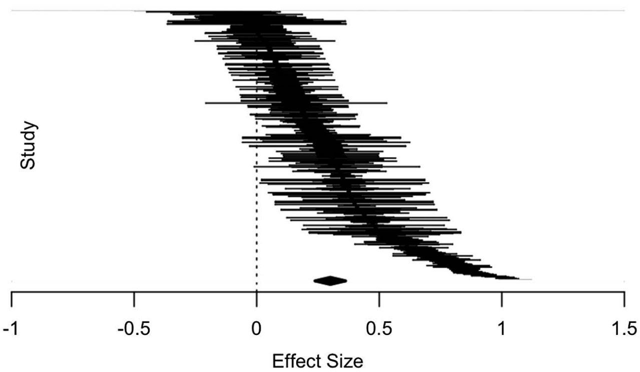
To explore the identifiability of different traits, we performed a series of tests. First, we explored whether the convergence of computer predictions and target ratings differed across the Big Five traits. Our results demonstrated a significant effect $F(4, 529) = 5.71$, $p \leq .001$. Then, we examined differences among the estimated effect sizes of the Big Five traits by including traits as categorical moderators, where effect sizes are grouped by each trait as a fixed effect in the multilevel model. Overall, Openness demonstrated the highest self-computer agreement ($\rho = 0.34$, 95% CI [0.28, 0.41]) followed by Conscientiousness ($\rho = 0.31$, 95% CI [0.24, 0.37]), Extraversion ($\rho = 0.30$, 95% CI [0.24, 0.37]), Agreeableness ($\rho = 0.28$, 95% CI

**Table 3**
*Meta-Analytic Analyses for the Computer Prediction Studies*

| Type of analysis | $k_s$ | $k_{es}$ | βρ | $p$ | [95% CI] | ρ | [95% PI] | $\tau_1^2$ | $\tau_2^2$ | $F$ | $Q_{residual}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Overall effect | 42 | 534 | | <.001 | [0.24, −0.36] | 0.30 | [−0.15, 0.75] | 0.01 | 0.04 | | 454858.11** |
| Big Five Traits | 42 | 534 | | | | | | 0.10 | 0.20 | 22.59** | 435365.79** |
| Openness (intercept)[a] | 41 | 106 | 0.34 | <.001 | [0.28, −0.41] | 0.34 | [−0.11, 0.79] | | | | |
| Conscientiousness | 41 | 106 | −0.03 | .048 | [−0.06, −0.00] | 0.31 | [−0.14, 0.76] | | | | |
| Extraversion | 42 | 108 | −0.04 | .023 | [−0.07, −0.01] | 0.30 | [−0.14, 0.75] | | | | |
| Agreeableness | 41 | 106 | −0.06 | <.001 | [−0.10, −0.03] | 0.28 | [−0.17, 0.72] | | | | |
| Neuroticism | 42 | 108 | −0.07 | <.001 | [−0.10, −0.04] | 0.27 | [−0.17, 0.72] | | | | |
| Platform | 42 | 534 | | | | | | 0.011 | 0.029 | 2.20* | 298236.96** |
| Multiple platforms (intercept)[a] | 2 | 60 | 0.79 | <.001 | [0.54, 1.03] | | [0.32, 1.25] | | | | |
| Spotify | 1 | 5 | −0.43 | .047 | [−0.86, −0.06] | 0.35 | [−0.18, 0.88] | | | | |
| X | 6 | 155 | −0.58 | <.001 | [−0.86, −0.30] | 0.21 | [−0.21, 0.63] | | | | |
| Facebook | 13 | 144 | −0.51 | <.001 | [−0.77, −0.24] | 0.28 | [−0.13, 0.69] | | | | |
| XING | 1 | 5 | −0.43 | .052 | [−0.86, 0.00] | 0.36 | [−0.18, 0.89] | | | | |
| Flickr | 2 | 10 | −0.35 | .049 | [−0.70, −0.00] | 0.43 | [−0.04, 0.90] | | | | |
| Sina Weibo | 5 | 60 | −0.45 | .002 | [−0.74, −0.16] | 0.34 | [−0.09, 0.76] | | | | |
| Money app | 1 | 5 | −0.61 | .005 | [−1.04, −0.18] | 0.17 | [−0.36, 0.70] | | | | |
| Personal computer | 1 | 5 | −0.33 | .142 | [−0.77, 0.11] | 0.46 | [−0.08, 1.00] | | | | |
| Video | 4 | 20 | −0.67 | <.001 | [−0.98, −0.37] | 0.11 | [−0.32, 0.55] | | | | |
| Smartphones | 5 | 45 | −0.46 | .002 | [−0.75, −0.17] | 0.32 | [−0.10, 0.75] | | | | |
| Servicelogs | 1 | 20 | −0.55 | .011 | [−0.97, −0.13] | 0.23 | [−0.29, 0.76] | | | | |
| Data type | 42 | 534 | | | | | | 0.011 | 0.038 | 1.44 | 216327.91** |
| Multiple data (intercept)[a] | 19 | 182 | 0.38 | <.001 | [0.29, 0.47] | | [−0.07, 0.82] | | | | |
| Language | 9 | 195 | −0.19 | .019 | [−0.35, −0.03] | 0.19 | [−0.26, 0.65] | | | | |
| Picture | 5 | 92 | −0.15 | .136 | [−0.35, 0.05] | 0.23 | [−0.24, 0.70] | | | | |
| Metadata | 3 | 15 | −0.18 | .150 | [−0.43, 0.07] | 0.20 | [−0.30, 0.69] | | | | |
| Likes | 1 | 5 | −0.04 | .848 | [−0.44, 0.36] | 0.34 | [−0.25, 0.93] | | | | |
| Smartphone | 5 | 45 | −0.06 | .567 | [−0.25, 0.14] | 0.32 | [−0.15, 0.79] | | | | |
| ML method | 42 | 534 | | | | | | 0.01 | 0.04 | 0.32 | 444352.35** |
| No validation (intercept)[a] | 5 | 22 | 0.23 | .018 | [0.04, 0.42] | | [−0.26, 0.72] | | | | |
| *k*-fold validation | 30 | 420 | 0.08 | .423 | [−0.12, 0.29] | 0.31 | [−0.15, 0.77] | | | | |
| Holdout | 7 | 92 | 0.08 | .533 | [−0.17, 0.32] | 0.31 | [−0.17, 0.78] | | | | |
| Type of publication | 42 | 534 | | | | | | 0.01 | 0.04 | 3.23 | 446130.823** |
| Journal (intercept)[a] | 17 | 197 | 0.23 | <.001 | [0.14, 0.33] | | [−0.21, 0.68] | | | | |
| Conference | 25 | 337 | 0.11 | .073 | [−0.01, 0.24] | 0.35 | [−0.09, 0.79] | | | | |
| Publication year | 42 | 534 | | | | | | 0.01 | 0.04 | 2.877 | 378875.96** |
| Intercept | | | 0.40 | <.001 | [0.27, 0.53] | | [0.27, 0.53] | | | | |
| Slope | | | −0.02 | .090 | [−0.04, 0.00] | | [−0.13, 0.09] | | | | |

*Note.* $k_s$ = number of studies; $k_{es}$ = number of effect sizes; βρ = regression coefficients; 95% CI = 95% confidence interval; ρ = corrected point estimate; 95% PI = 95% prediction interval; $\tau_1^2$ and $\tau_2^2$ = within- and between-study variance, respectively; ML = machine learning.
[a] Reference group for moderator analyses.
* $p < .01$. ** $p < .001$.

**Figure 3**
*Caterpillar Plot of the 534 Effect Sizes Included in the Meta-Analysis*




[0.21, 0.34]), and Neuroticism (ρ = 0.27, 95% CI [0.21, 0.34]). Further, pairwise contrasts (all of which were Bonferroni corrected) revealed several significant differences. Specifically, Conscientiousness can be more effectively identified than Agreeableness (contrast = 0.03, 95% CI [0.00, 0.06], $p = 0.05$) and Neuroticism (contrast = −0.03; 95% CI [−0.07, −0.03], $p = 0.03$). Openness can be more accurately predicted than Conscientiousness (contrast = 0.03; 95% CI [0.00, 0.06], $p = 0.05$), Agreeableness (contrast = 0.06, 95% CI [0.03, 0.10], $p < .001$), Extraversion (contrast = 0.04; 95% CI [0.01, 0.07], $p = 0.02$), and Neuroticism (contrast = 0.07, 95% CI [0.04, 0.10], $p < .001$). Finally, the 95% prediction intervals were wide, indicating a high level of heterogeneity for each trait (see Table 3).

### Publication Bias

Publication bias was first assessed by examining a funnel plot of effect sizes plotted against the inverse of the standard errors (see Figure 4). The funnel plot demonstrated an asymmetrical distribution of effect sizes. Specifically, there were some studies with smaller sample sizes, that in turn have higher standard errors and lower precision, that tended to produce smaller or negative effect sizes compared to those with larger sample sizes and greater precision. Moreover, the modified Egger's regression test demonstrated a significant effect of precision (β = 0.97, 95%CI [0.79, 1.14], $p < .001$). These findings suggest that studies with small sample sizes, that in turn may produce small effect sizes, may not have been published. It is worth noting that these findings do not confirm that publication bias exists, particularly given the substantial heterogeneity present in the results. As such, these results may be due to heterogeneity or other study quality effects (Ioannidis, 2005; Ioannidis et al., 1998; Terrin et al., 2003).

### Moderator Analyses

To explore sources of heterogeneity, we performed a series of moderator tests that examined different types of data, study, and methodological characteristics (see Table 3). Then, all significant moderators were entered into a multivariate model to determine sources of heterogeneity when all moderators are examined simultaneously (see Table 4). In both tables, we report βρ denoting the regression coefficients, 95% confidence intervals, and $p$ values for βρ, point estimates (ρ), and their prediction intervals, and heterogeneity statistics. We note that these analyses are exploratory due to the vast types of digital data, platforms, and approaches reported (and that exist but are not, or not yet, reported). As such, some of the moderators analyzed comprise small samples (e.g., where $k_s = 1$ or 2), and effect sizes (e.g., where $k_{es} = 5$ or 10). These analyses should therefore be interpreted with this in mind and treated with caution as analyses based on small $k_s$ may be less robust. We outline these findings below.

**Platform.** Significant moderation for platform and personality prediction was demonstrated for Spotify, X, Facebook, Flickr, Sina Weibo, Money App, Video, Smartphones, and Service Logs. There was no evidence of moderation by Xing, or Personal Computer. All significant moderators demonstrated a smaller effect size than the Multiple Platforms reference group. The 95% prediction intervals indicate heterogeneity across each platform.

**Data Type.** Our analysis explored whether language, pictures, metadata, likes, and smartphone data moderated the prediction of personality from digital data (multiple data types was the reference category). There was no evidence of moderation across these categories. The 95% prediction intervals indicate heterogeneity across each data type.

**Machine Learning Method.** There was no evidence of moderation across type of machine learning method used (i.e., k-fold validation, and holdout with no validation as the reference category). The 95% prediction intervals indicate heterogeneity across each method.

**Type of Publication.** Significant moderation for type of publication was demonstrated, where conference proceedings were compared to journal publications (reference category). Journal publications demonstrated a significant and larger effect size than conference proceedings: βρ = 0.23, $p < .001$ and βρ = 0.11, $p = .073$, respectively. The 95% prediction intervals indicate heterogeneity across each type of publication.

**Publication Year.** There was no evidence of moderation for publication year.

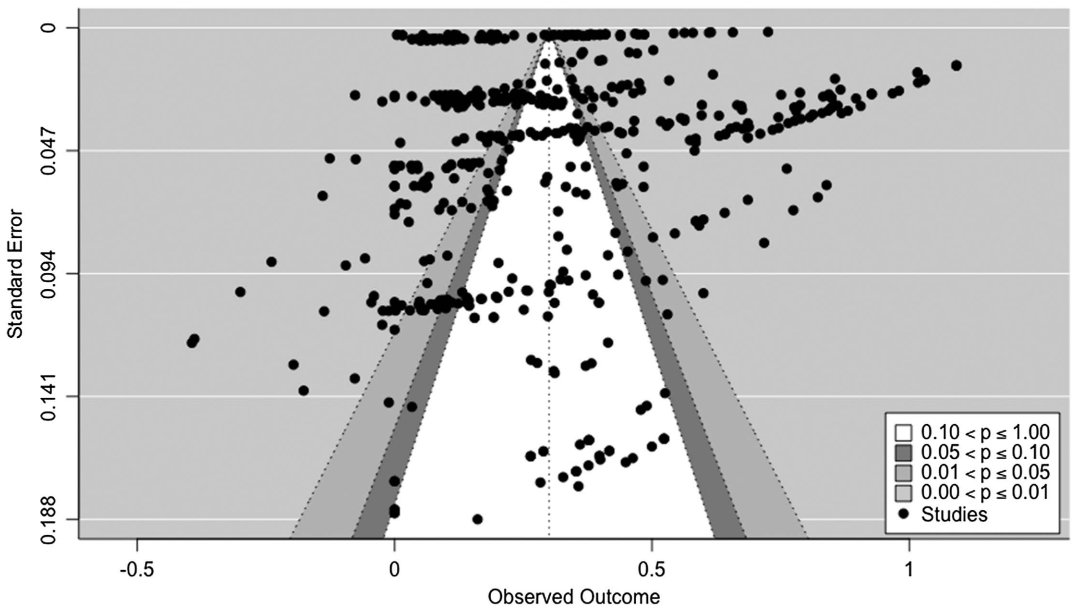
**Multivariate Model.** Following the aforementioned analyses, all significant moderators were included in a final model. The multivariate model demonstrated significant moderation, $F(10, 523) = 1.878$, $p = .046$ (see Table 4).

### Discussion

Overall, our meta-analysis demonstrates moderate convergent validity for computer-based personality predictions. Akin to our findings from the human perception meta-analyses, this evidence offers further support that digital traces reflect people's personalities, and that computer algorithms can be trained to predict personality using digital data. Table 5 provides a summary of the main effect sizes from each study. Similar to the human perception meta-analyses, convergent validity was highest for Openness and lowest for Neuroticism. However, the mean effect sizes overall were lower for each trait. There are several possible explanations for this finding.[7] One is that this is likely due to the methodological

[7] It is also worth noting that the Schmidt–Hunter approach uses sample sizes as weights, whereas the multilevel meta-analysis uses sampling variances. Such weighting schemes can also produce differences in mean effect size estimates. The differences between the effect sizes must therefore be interpreted with caution.

**Figure 4**
*Funnel Plot to Assess Publication Bias*




approaches and reporting conventions adopted in the computer prediction studies. For instance, researchers typically experiment with different algorithms and combinations of data until an optimal approach is achieved. If researchers are transparent about all of the approaches they have taken, then their reported findings will comprise a range of effect sizes, that will lower the mean effect size in a meta-analysis. Indeed, this is reflected in the within-study variance ($I^2$ = 22.21%) at level two of the multilevel analysis.

Another explanation is that algorithms are not as good at identifying traits than humans at present. Researchers' focus on optimization often occurs without consideration of why certain types of data are more indicative of personality than others meaning that researchers do not make decisions or use meaningful insights in creating algorithms that improve personality predictions. As such, this could stifle progress as researchers may make similar decisions or mistakes in pursuit of improving convergent validity over the underlying relationships between digital data and personality that may be informative. Indeed, although the Youyou et al.'s (2015) study demonstrated that algorithms could predict personality better than humans, no other research has made efforts to compare human and computer personality prediction or to disentangle the factors that may make one better than the other (or how these approaches may provide different information about personality; we discuss this further in the General Discussion).

Our findings also indicate that computer models are better at predicting some traits than others. Openness can be more effectively identified than all other traits, and Conscientiousness can be more effectively identified than Agreeableness and Neuroticism. Theoretical accounts of personality would therefore argue that Openness and Conscientiousness in this context are considered

**Table 4**
*Results of the Multivariate Model for the Computer Prediction Studies*

| Moderator | $k_s$ | $k_{es}$ | βρ | $t$ | [95% CI] | ρ | [95% PI] | $\tau_1^2$ | $\tau_2^2$ | $Q_{residual}$ | $I^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Multivariate model | 42 | 534 | | | | | | 0.01 | 0.03 | 297449.26** | 1.878* |
| Intercept | | | 0.55 | 5.16** | [0.34, 0.76] | | | | | | |
| Spotify | 1 | 5 | −0.20 | −0.94 | [−0.62, 0.22] | 0.35 | [−0.20, 0.90] | | | | |
| X | 6 | 155 | −0.39 | −3.24* | [−0.63, −0.15] | 0.16 | [−0.28, 0.61] | | | | |
| Facebook | 13 | 144 | −0.32 | −2.98* | [−0.53, −0.11] | 0.23 | [−0.20, 0.67] | | | | |
| Flickr | 2 | 10 | −0.18 | −1.13 | [−0.51, 0.14] | 0.37 | [−0.14, 0.87] | | | | |
| Sina Weibo | 5 | 60 | −0.26 | −2.04* | [−0.50, −0.01] | 0.29 | [−0.15, 0.75] | | | | |
| Money App | 1 | 5 | −0.38 | −1.76 | [−0.80, 0.04] | 0.17 | [−0.38, 0.73] | | | | |
| Videos | 4 | 20 | −0.44 | −3.09* | [−0.72, −0.16] | 0.11 | [−0.34, 0.56] | | | | |
| Smartphones | 5 | 45 | −0.27 | −2.14* | [−0.79, 0.02] | 0.28 | [−0.17, 0.73] | | | | |
| Service logs | 1 | 20 | −0.38 | −1.85 | [−0.79, 0.02] | 0.17 | [−0.39, 0.73] | | | | |
| Conference | 25 | 337 | 0.06 | 0.93 | [−0.07, 0.20] | 0.62 | [0.16, 1.07] | | | | |

*Note.* $k_s$ = number of studies; $k_{es}$ = number of effect sizes; βρ = regression coefficients; 95% CI = 95% confidence interval; ρ = corrected point estimate; 95% PI = 95% prediction interval; $\tau_1^2$ and $\tau_2^2$ = within- and between-study variance, respectively
* $p < .01$. ** $p < .001$.

"good traits" (e.g., Funder, 1995, 1999), as they manifest more clearly in digital data than the other traits. For Openness, this may be due to targets leaving indicators of their varied interests, creativity, emotions, etc. (e.g., liking different groups on Facebook, creating and posting artistic photographs). For Conscientiousness, algorithms may be able to detect patterns that reflect reliability and consistency (e.g., responding to social media posts or text messages in a timely manner, using language that is polite and friendly). It may also be the case that Agreeableness and Neuroticism are difficult to identify because they are internalized and are therefore more difficult to detect (similar to our findings for human perception). Alternatively, it could be that the current approaches have simply not found types of data, aggregations of data, and patterns that relate to these traits in more apparent ways yet.

Although human perception generally appears to be better than computer prediction across all traits, variances exist in the ranking of each trait's identifiability (see Table 5). Notably, Openness stands out as the most identifiable trait for both humans and computers, but the disparity in identifiability between the two is more pronounced than any other trait. This suggests that the cues humans rely on to discern Openness may be more revealing than the data computers utilize. Alternatively, humans appear to be particularly good at recognizing Extraversion, whereas computers are only middling in this regard. This suggests that the online cues humans utilize to detect Extraversion might by highly visible and minimally evaluative, consistent with theories of human perception offline (e.g., Beer & Vazire, 2017; John & Robins, 1993). In contrast, the digital data driving computer predictions might not align with these explanations. Given that computer predictions are developed from a bottom-up analysis of digital data, the concepts of high observability or evaluativeness might not be as applicable. For instance, although human perception studies often focus on social media profiles, computer analyses might delve into data that are not a primary reference for human judgment, such as smartphone logs, or tweet counts, or data that humans find difficult to interpret, like network metrics. Therefore, merely having access to a broader spectrum of data might not necessarily enhance personality inferences beyond human observations using current methodologies. This reasoning might also explain the patterns observed with Agreeableness, a trait typically perceived as moderately observable and evaluative by humans.

The identifiability of Neuroticism and Conscientiousness is low across both human perception and computer prediction. This aligns with research suggesting that individuals high in Neuroticism tend to be discreet about the information they disclose (Loiacono, 2015), and may portray idealized or false self-images online (e.g., Michikyan et al., 2014; Twomey & O'Reilly, 2017). Similarly, conscientious individuals tend to engage less frequently in online posts and comments (Lee et al., 2014; Ryan & Xenos, 2011) and exhibit cautious self-presentation (Seidman, 2013). Consequently, identifying Neuroticism and Conscientiousness is challenging, whether assessed by humans or computers. Although it seems plausible that both traits may be more identifiable from behavioral residue (i.e., the "hidden" traces of data rather than observable cues), current approaches have not fully exploited this possibility. Our current understanding therefore remains constrained when it comes to contextualizing how various data types relate to each trait across human perception and computer prediction. Future research should therefore seek explanations for the relationships that underpin them. Our meta-analysis reveals that most of the variability among the included effect sizes is linked to study level differences ($I^2$ = 77.70%). The presence of substantial heterogeneity is not surprising given the diversity of media, platforms, devices, types of data, and methods included in our analysis and is often expected in such circumstances (Higgins, 2008). Our analyses demonstrated moderating effects among different platforms, specifically X, Facebook, Sina Weibo, Videos, and Smartphones. Such differences may be explained by different patterns of usage among individuals, for instance, using X to follow the news and post public communications versus sharing information amongst a closed network of friends on Facebook. Sina Weibo is also likely to reflect cultural differences between users given that its user base is Chinese and other platforms are not available in China. Videos offer vastly different types of data to other types of media/platforms, as individuals can display/edit aspects of themselves that are not captured by other digital data such as tone of voice, body language, accent. Further, smartphones offer diverse forms of behavioral data traces that individuals leave behind unconsciously (e.g., GPS co-ordinates, number of screen locks, times of phone calls) that therefore link to different traits. These findings therefore highlight the need to investigate the underlying reasons for this heterogeneity in future research. We discuss these ideas in the General Discussion below.

**Table 5**

*Summary of the Main Effect Sizes for Each Trait Across Study 1 and Study 2*

| | Human study | | | Computer study | | | |
|---|---|---|---|---|---|---|---|
| Trait | $k_s$ | $\rho^a$ | [95% CI] | $k_s$ | $k_{es}$ | $\rho^a$ | [95% CI] |
| Openness | 29 | 0.57 | [0.53, 0.61] | 41 | 534 | 0.34 | [0.28, 0.41] |
| Conscientiousness | 29 | 0.39 | [0.36, 0.42] | 41 | 534 | 0.31 | [0.21, 0.37] |
| Extraversion | 30 | 0.52 | [0.48, 0.55] | 42 | 534 | 0.30 | [0.24, 0.37] |
| Agreeableness | 29 | 0.46 | [0.42, 0.51] | 41 | 534 | 0.28 | [0.21, 0.34] |
| Neuroticism | 28 | 0.38 | [0.35, 0.41] | 42 | 534 | 0.27 | [0.21, 0.34] |

*Note.* $k_s$ = number of studies; $k_{es}$ = number of effect sizes; ρ = corrected point estimate; 95% CI = 95% confidence interval.

[a] Effect sizes must be interpreted with caution; the human perception and computer studies were performed using different methods (Schmidt–Hunter and multilevel modeling, respectively) data sets, and analytical approaches that may not be fully comparable.

## General Discussion

Overall, our findings demonstrate that both humans and computer algorithms can infer personality from digital data with moderate success. Convergent validity is higher for Openness and lower for Neuroticism across both. Although these findings broadly align with theoretical models that posit that certain traits are "good" (i.e., they are more visible and easily detectable than others, Funder, 1995), our review demonstrates that we know little about which types of digital data (and associated platforms/technologies) are indicative of personality and why such relationships may exist (i.e., their content validity). These limitations are due to methodological and disciplinary conventions that prevail across psychology and computer science research. Our synthesis has enabled us to systematically analyze these approaches, and to identify specific limitations and opportunities that could help to address these gaps in future research. We encourage researchers to adopt complementary approaches that incorporate the strengths of both human perception and computer prediction and embrace new approaches that seek to improve construct validity as well as convergent validity. This will enable researchers to gain new understanding of personality and its theoretical underpinnings that could inform many practical applications from recruitment to medical diagnosis. These insights are discussed below.[8]

### Reconsidering Approaches to Sampling and Data Collection

#### *Sampling*

Approaches toward data collection and sampling could be substantially improved in future research. For instance, human perception research usually comprises psychology student volunteers or research assistants, who donate their digital data for use in the studies or participate as observers. This process has prevailed, despite calls for researchers to diversify their recruitment of samples across psychological science in recent years (Henrich et al., 2010a, 2010b; Johnson, 2021). Researchers' overreliance on Western, educated, industrialized, rich, and democratic (Henrich et al., 2010b) populations is therefore likely to impact *convergent validity* as observers may be better at assessing targets possessing similar demographic attributes, that are from similar cultural backgrounds and that use similar technologies and social media platforms to themselves. Observers from different backgrounds or cultures to the targets they assess may also utilize digital data in different ways to formulate judgments that may affect the identifiability of traits, and the predictability of cues.

More broadly, it is possible that these issues may impact measurement invariance, the extent to which the underlying constructs are measured in the same way across different groups (e.g., Meredith, 1993; Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). Measurement invariance has been shown to be problematic in personality research, for instance, Dong and Dumas (2020) found that studies of measurement invariance across cultures was lower than measurement invariance across age-groups or gender. However, at present we do not know if, or to what extent, measurement invariance may be inherent in current research.

Another related problem is that participants are unlikely to be malicious, fraudulent, or deceptive users—presumably such individuals are unlikely to donate their profiles to psychological research. Thus, although there is some evidence to suggest that people attempt to present themselves honestly (N. Ellison et al., 2006), or their "real" selves online (Back et al., 2010), such evidence is also derived from these limited samples. Indeed, although gaining such data from dishonest or malicious users would be challenging and pose ethical challenges in terms of collecting and analyzing their data (see "Privacy, Trust, and Ethics" below), research into such populations could greatly benefit human perception research and would open new avenues of theoretical inquiry in gaining understanding of how cues constitute "good information" online (Funder, 1995, 1999). Such work could also have significant societal impact in terms of addressing issues such as identity theft (e.g., Alharbi et al., 2021; Tsikerdekis & Zeadally, 2014), grooming (Whittle et al., 2013), and online dating (e.g., Coluccia et al., 2020; Whitty, 2018).

In contrast, although computational approaches are able to overcome some of these sampling challenges by harvesting a breadth of demographics at scale, and by analyzing individuals from different cultures, many studies in our set comprised relatively small samples that ranged from a few hundred to a few thousand, and only a few analyzed larger data sets comprising tens or hundreds of thousands of individuals (see Liu, Preot, & Ungar, 2016; Schwartz et al., 2013; Xue et al., 2018; Zhang et al., 2018). We also know little about the specific demographics or cultures that comprised those data sets as minimal detail was described in the reports (and often no detail was disclosed whatsoever). Thus, although computer prediction studies may be better placed to investigate and address issues relating to measurement invariance, akin to human perception research we do not yet know the extent to which such issues may prevail.

Further, algorithms are trained and tested within the same sample meaning that they will be optimized to predict personality using those demographics. Current findings may therefore be skewed in favor of particular demographics or populations that can impact the generalizability of the algorithms' performance, as the same results may not be achieved on different samples. Other factors can also impact the representativeness of a sample—for instance, such data will not reflect the demographics/populations of users that do not have/use (or cannot use) social media, smartphones, or other devices. Likewise, occasional, or low users of technologies leave different types of data to heavy users, and those who choose not to use social media or other devices will not be captured whatsoever. If such issues are not addressed, then subsequent implementations of personality assessments could exacerbate social inequalities and the digital divide. Researchers should therefore make efforts to diversity samples and to reflect on (and communicate) any limitations so that underrepresented or difficult to access populations are not neglected.

#### *Data Collection*

Human perception studies are typically performed in laboratory-based environments where targets' information is presented to observers as a single "snapshot" of the target at a particular moment in

[8] We also note that while numerous studies were omitted from the meta-analysis in Study 2 due to the metrics reported, the platforms, types of data, and methodological approaches applied did not differ from the studies that were. Thus, our insights reported below are derived from all studies across the human perception and computer prediction literature that met our main inclusion criteria, and that are presented in Supplemental Files 2 and 3.

time (e.g., forming judgments from a set of tweets, Qiu et al., 2012), or from browsing their profiles for a few minutes (Back et al., 2010; Kluemper et al., 2012). While these approaches benefit from improved ecological validity that early studies of impression formation (that relied on material constructed by the researchers, e.g., videos of targets being interviewed Gangestad et al., 1992, or mock jury sessions, Scherer, 1972), they do not reflect how individuals use digital sources to form impressions in reality. For instance, making judgments likely consists of consulting multiple sources (e.g., browsing a person's Instagram account, X profile, workplace staff page) and browsing the content over different periods of time (i.e., posts made this year, and then 5 years ago) as opposed to screenshots that do not allow this flexibility and exploration.

In contrast, computer algorithms can gather, process and analyze data "in the wild" dating back years and can analyze the entire set of data that are available for a person. Laboratory setups are time consuming and costly, and although it would be impossible to ask observers to comprehend the type/volume of data that algorithms can process, improvements could be made to the way that studies are conducted. For instance, encouraging observers to spend time considering different elements of the information presented (e.g., photographs, language, network information), could provoke them to formulate judgments in a more focused and meaningful way, enabling researchers to explore the content validity of different types of data and to explore psychological theories/explanations of how and why observers connect digital data to personality traits. Similarly, broadening the sources used to formulate judgments (e.g., social media profiles combined with smartphone usage data, energy consumption, movement patterns, fitness data, and so forth) could help observers to build richer, more informed judgments by enabling them to draw upon cues that reflect targets' behavior across different contexts and circumstances. Researchers could use these approaches to further test and extend personality models and theories. For instance, in line with Vazire (2010) self-other knowledge asymmetry (SOKA) model, researchers could explore whether asymmetries exist in judgments made from different media (e.g., how judgements based on a target's smartphone records compare to their X profile), or whether data from multiple sources reveals new imbalances or unknown aspects of targets' personalities.

Although computational approaches have capitalized on the use of bigger data sets and analysis of data over longer periods of time, existing work has also predominantly focused on data collected from one platform or device. Computer algorithms may therefore be able to gain valuable insights into personality by drawing upon data from multiple sources (i.e., via *cross media-* or *cross domain-learning*; Joshi et al., 2012; Zhuang et al., 2010). For example, a computer model could be trained using smartphone usage data and X data, and the model could then be used to predict personality from individuals' fitness tracker data. Thus, training computer models by combining data from different platforms and devices could improve the personality predictions or could be used in situations where representative data are not available. In turn, this could also help to address the aforementioned issues relating to the generalizability of findings. Future work could therefore delve deeper into this opportunity to investigate ways to improve predictions and to discover new behavioral patterns.

Digital data can also be used as experience sampling methods (ESM) or as ecological momentary assessments (EMA). Traditionally, ESM and EMA have relied upon self-reports completed by subjects who are prompted to report on their behavior, experiences, thoughts, beliefs or feelings repeatedly over a period of time (Larson & Csikszentmihalyi, 2014; Shiffman et al., 2008). Thus, digital data are potentially rich sources of experience sampling because they can be collected unobtrusively as people navigate their daily lives (e.g., use their smartphones, cookers, washing machines, televisions) and generate data across the many different contexts they occupy or inhabit. Thus, the adoption of what is known as *personality sensing* (Harari et al., 2020) methodologies could broaden investigations to encompass the examination of personality states in addition to traits, as well as a means to predict future behavior (Conner et al., 2009; Fleeson & Gallagher, 2009). A number of recent studies have started to explore this potential in related areas, such as how experience sampling of mobility patterns (obtained via GPS tracking) relates to subjective well-being (Müller et al., 2020), how people's movements (obtained via GPS sensors) relate to personality traits, states, and momentary expressions (Matz & Harari, 2021), and how personality relates to sleep patterns (obtained via smartphone sensing data, Schoedel et al., 2020).

Future research could also exploit these opportunities to study the relationship between individuals' thoughts, feelings, attitudes and motivations and digital data—areas that (to our knowledge) have been largely overlooked in personality research within this context. New forms of ESM/EMA where data are collected passively (i.e., data generated automatically via sensors, GPS, smartphone logs, and so forth) could be combined with active data collection (or "active ESM"; Shiffman et al., 2008) where subjects report on their feelings, mood, etc. could help to provide understanding of how these cognitive states manifest in digital data, and indeed how these states could relate to (or help to further develop) existing psychological theory. Smartphones provide an ideal means of recording such information, as they are widely used by the majority of the population (e.g., Ellis, 2020; Wilcockson et al., 2018) and therefore capture a breadth of a user's activity, from call logs and application data, to GPS tracking and sensor data, as well as a means for individuals to record self-report data periodically. Several applications have recently been developed that could support future work. These include, Unforgettable.me (Dennis et al., 2019) that collects image, GPS, accelerometer, and audio data, SurveySignal (Hofmann & Patel, 2015) that integrates text messages as signals or reminders to complete surveys, and PEG LOG (Geyer et al., 2019), that records users' location data.

## Convergent and Construct Validity

Current approaches in human perception and computer prediction are entirely focused on convergent validity, or the "accuracy" of self-other or self-computer personality assessments. In human perception studies, researchers typically investigate whether convergent validity between observers and targets (or consensus amongst observers) exists. As demonstrated by our meta-analyses, effect sizes vary, and often appear on the lower side. While such magnitudes may be perceived by some to be weak or to provide little insight (e.g., see Mischel, 1968), it is important to consider their substantive significance within the broader context of personality research. Such effect sizes can be very useful and help to inform application and theoretical development (see Funder & Ozer, 2019). Moreover, although it can be argued that such approaches tend to view targets' self-reports as "the truth" about their traits, assessing the convergent validity of self-other agreements offers a means for researchers to experiment with different approaches and to test whether observers' assessments improve (akin to training computer algorithms). In a similar vein, researchers sometimes explore which cues (i.e., types of digital data) are utilized by observers or are valid indicators of personality (Back et al., 2008; Darbyshire et al., 2016; Stopfer et al., 2014) to provide additional context to observers' impression formation. Although such findings align with models that argue that information must be "good" to make accurate judgments (Funder, 1995, 2012) and that observers must be able to utilize valid cues (Brunswik, 1956), no attempts are made to understand why different types of digital data, or the types of technologies that generate them, may relate to specific traits.

Researchers may therefore want to consider the impact that training or incentivizing individuals to complete assessments truthfully and to the best of their ability may have on the effectiveness of personality assessments. Likewise, making efforts to improve individuals' abilities, and exploring how and why observers make judgements using specific cues could help to move research beyond asking whether observers can infer personality from digital data to instead focus on the practical value that potential improvements could have. For instance, training individuals to look out for cues of a fraudulent or deceptive person could help to combat phishing and online dating scams. Similarly, educating more diverse and vulnerable populations such as children, the elderly and neurodiverse individuals will have important implications for their understanding of online interactions and for their safety online.

In contrast, although computer prediction studies do not tend to be informed by the aforementioned frameworks, researchers' motives are similar in that their primary aim is to achieve some level of convergent validity, based on whether this is possible using different types (or combinations) of digital data. Convergent validity is important for assessing how effectively an algorithm can predict personality using the data it has been trained with. However, researchers tend to include all data that are useful in a model when attempting to find an optimal solution. This can affect the discriminant validity of the tests because digital data can often be related to multiple traits, for instance, in an investigation of music listening behavior on Spotify, Anderson et al. (2021) found that "acousticness entropy" positively correlated with both Openness and Agreeableness, and that Brazilian Music correlated with Openness, Agreeableness, and Emotional Stability. It can also affect content validity as new relationships can be identified (e.g., Anderson et al., 2021, found that "mood entropy," "mechanism entropy," and "organism entropy" were all positively correlated with Openness), yet can be difficult to make sense of. Thus, without efforts to explain such findings, it is impossible to ascertain whether such results are opening new theoretical areas of investigation, or whether they are spurious and would not replicate in another study.

Such issues have been broadly acknowledged in the field of psychology for some time (e.g., Bornstein, 2009; D. T. Campbell & Fiske, 1959) and despite calls for researchers to address these problems (Bleidorn & Hopwood, 2019), existing work has neglected to do so. Thus, researchers' sole focus on convergent validity is problematic because the resultant findings offer no insight into why digital data relate to personality, meaning that current work cannot test or offer insights into psychological theory. That said, computational approaches are extremely well-placed to discover patterns in data from the ground up to develop new frameworks, theories, and models.[9] These findings could challenge existing measures of personality and remove the need for individuals to complete self-report questionnaires. Although their approach did not seek to develop or contribute toward theory, Gerlach et al. (2018) demonstrated this possibility by conducting a data-driven analysis of a large data set comprising 1.5 million individuals. Their findings provided evidence for four distinct personality types, that contrast with existing approaches that have categorized personality into five or six traits (i.e., the FFM; McCrae & Costa, 1987 or the Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, Openness to Experience model; Ashton et al., 2004). Future research could therefore adopt these approaches by combining data-driven analytics with theory development that could result in new definitions and models of personality altogether. For example, after identifying patterns in data, researchers could theorize about why such relationships may exist and perform experiments to test their hypotheses.

Consider Vazire's SOKA model for instance. Vazire (2010) posits that an asymmetry exists between what the self knows and what another person knows about that individual. As such, internalized traits such as Neuroticism are better known to the self than others (because individuals typically hide/suppress their anxieties making them difficult to perceive), whereas traits that are highly desirable or undesirable can be better perceived by others (e.g., irritability) because individuals often distort their self-perceptions (i.e., they would not like to consider themselves as petulant). Alternatively, externalized traits such as Extraversion are known well by both individual/observer as they are outwardly displayed and visible (e.g., sociability). Similar to a Johari window (Luft & Ingham, 1955), the SOKA model contains four quadrants, each depicting which aspects of personality are best known to the self and those that are best known to others (specifically the "open arena" [where the self and others know an individual's traits], the "blind arena" [where only others know the individual's traits], the "hidden arena" [where an individual's traits are known to themselves but not others'], and the "unknown arena" [where an individual's traits are unknown to themselves and others], Vazire, 2010).

Computational methods could further our understanding of these asymmetries or even change these imbalances because the extent to which traits are observable or evaluative are not equivalent to human/offline personality perception. Digital data, particularly in the form of behavioral residue, are likely to convey objective information (that are removed from intentions to present oneself in a particular way) and if such data are indicative of certain personality traits, then it

[9] These approaches are notably similar to the development of early personality scales such as the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) in the 1940s, and the subsequent development of the California Psychological Inventory (CPI; Gough, 2012) in the 1950s. The MMPI was initially developed to improve poorly developed and tested assessments of psychopathologies (Loevinger et al., 1953; Thorndike & Stein, 1937). By employing an empirical criterion keying approach, items were selected for their ability to distinguish between groups of people with different psychiatric disorders. Alternatively, the CPI was designed to assess nonclinical, or everyday "folk-concepts" relating to people's behavior. Although both the MMPI and CPI were initially successful in distinguishing between people with different conditions and behaviors, problems subsequently arose over the poor content validity and discriminant validity of the items. As a result, these issues prompted revised versions of the scales (e.g., the MMPI-2, MMPI-A; Butcher & Williams, 1992), MMPI-A-RF; and the CPI; Gough & Bradley, 1996), and influenced the design of psychological assessments over the following years (Bleidorn & Hopwood, 2019).

may be possible for an algorithm to "see" a trait that was previously difficult to detect (i.e., improve *detectability*). For example, creativity may be related to a person's betweenness centrality score (a measure of how closely people in a network are connected), if for instance, certain others within someone's network were highly creative. Such information may never be visible to a human perceiver yet could be readily detectable by an algorithm. Researchers could therefore account for/explore these possibilities to address potential asymmetries when creating algorithms. Further, algorithms may be able to explore the "unknown arena" within Vazire's SOKA model. By analyzing patterns of individuals' data bottom-up, new (previously unknown) aspects of their behavior may emerge and thus reveal people's blind spots in terms of how their self-perceptions (and others that observe them) are created and distorted.

Incorporating theoretical models such as Funder's RAM or Vazire's SOKA model into the design and implementation of studies could also help researchers who typically adopt data-driven approaches to think about the benefits that theory could offer. For instance, considering what may make particular data *relevant, available or detectable* could be a starting point for selecting data and hypothesizing *why* they might relate to personality. Testing these hypotheses could help to *explain* the resultant findings, and these findings could then inform future hypotheses or theory development in subsequent research. Research could also continue to exploit the benefits of data-driven principles. Selecting data for which there is no intuitive reasoning for testing could be a way to discover new patterns and relationships. It may also be the case that such data, in combination with data that has known relationships with personality boosts the predictive performance of algorithms and helps to further develop hypotheses/theory. Testing different combinations of data (known as *feature engineering*) is a crucial part of the modeling process in machine learning and incorporating theory into this process could bring great advances to the field moving forward. For some good overviews of this process, see Stachl, Pargent, et al. (2020) and Bleidorn and Hopwood (2019).

Researchers could also make concerted efforts to develop and test theory in this way through the reanalysis of existing data sets. Current approaches miss out on the vast possibilities to test the validity of various approaches and to determine their reproducibility. Open Science principles have gained prominence in recent years, and as a result, researchers are encouraged to transparently report methods, openly share research, data and code, etc. (e.g., Foster & Deardorff, 2017; McKiernan et al., 2016; Open Science Collaboration, 2015; Pineau et al., 2021) in order to better support reproducibility and replicability. These principles also extend to *computational reproducibility*, that focuses on aspects such as the documentation of code, the software environment, and version control that may also impact the replicability and reproducibility of findings (e.g., Crüwell et al., 2023; Stodden et al., 2018).[10] In the context of personality research, encouraging researchers to adopt Open Science practices, and to prioritize the creation, testing, and reporting of reproducible solutions could therefore revolutionize the way personality is assessed and understood, irrespective of the discipline it is reported in.

Another consideration is that the different approaches adopted across human perception and computer prediction could provide different information about personality. In their discussion of how machine learning could benefit from adopting the principles of construct validation, Bleidorn and Hopwood (2019) argued that research would benefit from treating different tests as complementary rather than competitive. Thus, rather than trying to determine "which test works best?" researchers could ask, "what do different tests tell us about an underlying construct?" (Bleidorn & Hopwood, 2019, p.198). They argue that poor content validity, overlapping constructs, and natural co-occurrence could result in convergence between different traits. Such convergence could suggest a need for modifications in the tests. Alternatively, it might indicate that the tests are capturing co-variance in nature, that could provide interesting information about the trait. Researchers could also extend these ideas to incorporate findings relating to human perception. For instance, insights gained from different types of observers (or observers that have different levels of knowledge, training, abilities, etc. in rating targets' personalities) may offer insights into personality that computer algorithms cannot detect. Researchers could therefore take insights from different approaches to inform others (e.g., use human perception ratings to inform the design of algorithms), or to evaluate human and computer ratings in tandem to build a more holistic profile of a person.

## Facets and Nuances

Big data and machine learning also offer the potential to explore how different data relate to personality facets, the subset of characteristics that comprise each trait (e.g., Agreeableness comprises facets such as sympathy, modesty, cooperation, and altruism). For instance, it may be the case that a particular type of data reflects the sympathy facet, but not the altruism facet. Indeed, building on the ideas mentioned above, researchers may therefore be able to explore whether and why digital data relate to certain facets to further understand content validity and to test psychological theory. To date, research has almost exclusively focused on domain-level predictions, bar a small number of exceptions (specifically Park et al., 2015 and Stachl, Pargent, et al., 2020). Presumably, this has been influenced by researchers' desires to obtain personality ratings that require little cost or effort, as reflected in the high number of studies that have used the shorter form personality scales, such as the Ten Item Personality Inventory or the Big Five Inventory, BFI-44. The studies that have examined the predictive potential of data in relation to facets, have so far demonstrated similar findings to domain-level predictions. That is, Park et al. (2015) found that Facebook status updates could predict the majority of facets, and Stachl, Pargent, et al. (2020) found that personality could be predicted at both the domain and facet level using smartphone behavioral data for all traits except Agreeableness.

Further, Hall and Matz (2020) recently expanded analyses beyond the facet level to encompass personality nuances that reflect the unique variance attributable to the items in a facet scale (McCrae, 2015; Mõttus et al., 2017). They found that nuance-based models (that used Facebook Likes) made small improvements in personality prediction when compared to domain-based models. Together, these studies show the benefits of expanding current methods to focus on finer grained analyses of personality and nuances. Such analyses

[10] Open Science practices are also increasingly being required as part of journals' submission policies, for example, see Crüwell et al. (2023) and Brysbaert et al. (2021). Thus, researchers would be well-advised not only to adopt these practices to improve reproducibility and replicability, but also to ensure that their work aligns with changes in journal/conference policies which will likely impact all disciplines and outlets eventually.

offer the potential to expand theoretical analyses of individuals' behavior, as well as to improve the external validity of the predictions. Future research would benefit from continuing to build on these approaches by integrating these analyses with data obtained from new technologies and devices (e.g., smartphones, wearables, IoT devices).

## Privacy, Trust, and Ethics

Predicting personality effectively has many practical implications—it can influence who we invite into our lives, impacting how effectively we work, the relationships we build, and how happy and healthy we feel. In a world where digital information is at our fingertips, it is inevitable that many of us will turn to the internet to help formulate life decisions involving other people. Thus, successfully making judgments from strangers' digital data and predicting personality computationally increases the potential to anticipate others' future behaviors, thoughts, and feelings. In many circumstances, the applications of computational methods are widely known and generally accepted, for instance, recommender systems that suggest films that people may like to watch, or products they might like to buy. Conversely, in other cases, such techniques are causing increasing concern about how digital data are accessed and analyzed, and what the consequences of these activities may entail. Perhaps the most illustrious example of this is the Facebook-Cambridge Analytica scandal in 2018, where the data analytics company Cambridge Analytica were accused of attempting to influence people's voting patterns in the 2016 U.S. Presidential Election, and the Vote Leave campaign in Britain's European Union referendum (Cadwalladr, 2018; Kitchgaessner, 2018). By illicitly harvesting data from an estimated 87 million individuals' Facebook accounts, Cambridge Analytica allegedly created a series of psychographically targeted advertisements that attempted to influence individuals' political preferences. Although we do not know the extent to which these efforts were truly effective in the case of Cambridge Analytica, the two key issues—the notion of mass-manipulation, and that of collecting individuals' data without their consent, provoked global outrage and calls for users to delete their accounts (Slawson, 2018; Timms & Heimans, 2018). In turn, related research on targeted advertising has demonstrated its potential effectiveness, for instance, Matz et al. (2017) found that microtargeted advertisements increased clicks and purchases by around 40% and 50%, respectively, compared to unpersonalized ads, and Zarouali et al. (2020) found that political ads that were congruent with individuals' personalities increased their intentions to vote.

The Cambridge Analytica scandal also highlighted that people do not fully understand the risks associated with their online interactions and are often unaware of how their data might be used (e.g., Bakir, 2020; Hinds et al., 2020; Larsson et al., 2021). This is further complicated by the fact that personal information can be inferred through implicit patterns in individuals' data, rather than via explicit expressions of their preferences (e.g., religion from Facebook likes), as well as from others within their networks (meaning they may "reveal" no information directly themselves). Recent research has demonstrated the relative ease by which information can be gleaned through individuals' profiles, for instance, a person's sexual orientation could be identified from their network connections (Garcia, 2017), facial images, (Wang & Kosinski, 2018) and tweets could be used to predict what an individual may say (or tweet about) in future (Bagrow et al., 2019). For psychologists, these same techniques raise critical questions around informed consent, proportionality in research, and the potential abuse of methods, applications, and findings. Although digital data may be "public," that does not necessarily mean that it is ethical to infer "hidden" characteristics to users who have not consented to such an action. Thus, the control of what personal information is disclosed and how data "spreads" through people's networks raises questions of where the responsibility for managing or protecting such information lies. For instance, should individuals be thinking about others' privacy when they use digital devices? And how should researchers and organizations ensure they obtain proper informed consent when using data that is inherently networked?

Current estimates predict that around 30 billion devices will be connected to each other by 2030 (Vailshery, 2022). As our interconnections continue to grow (with our devices as well as each other), the need to address these delicate ethical challenges will become increasingly pertinent. There are increasing demands for organizations and regulators to make their algorithms "transparent" to enable individuals to better protect their data (e.g., Buhmann et al., 2020). Likewise, researchers and organizations alike will need to consider transparency in terms of how they intend to use people's data, and in communicating what can be potentially "leaked" or inadvertently revealed through digital interactions (see Shaw et al., 2023, for guidelines on how to use and manage digital data ethically). The potential to analyze data in these new innovative ways could offer great insight to human behavior, with prosperous benefits to domains such as health care, finance, and criminal investigation. Further research needs to find ways that enable such aspects to flourish, while mitigating individuals' concerns, and potential risks to their private information.

## Limitations

There are several limitations that readers must consider when interpreting our findings. First, although we systematically and rigorously extracted data from both the human and computer sets of studies, the vastly different approaches adopted substantially limited what we could explore in our meta-analytic investigations. That is, the diversity of types of digital data used and the tendency to focus analyses on social media platforms such as Facebook, for instance, limited our moderator analyses, meaning that we could mostly describe results rather than draw meaningful conclusions from them. Second, the small number of studies available for each of the platforms, types of digital data used, etc. (particularly in the human set of studies) means our results could be subject to second-order sampling error, and that effect sizes will be more vulnerable to statistical artifacts. Third, the diversity of approaches and reporting conventions across the studies (both within and across disciplines) made it difficult to evaluate all methods, perspectives, analyses, etc. completely. Moreover, there are no formal mechanisms that readily support the evaluation of studies beyond the analysis of effect sizes in meta-analyses. For example, in many reviews and meta-analytic studies, it is possible to perform quality assessments using a set of predetermined scales or criteria such as the Newcastle-Ottawa Scale (Wells et al., 2000) or the JADAD Scale (Jadad et al., 1996). Although existing assessments cannot be readily applied to the reviewed studies, future research could therefore seek to develop scales or checklists to support such analyses. Fourth, our findings offer little conceptual or theoretical advancement in relation to personality in and of themselves, due to the disciplinary norms discussed earlier (i.e., prioritization of convergent validity, and the optimization of a computer algorithm's performance). We hope that our discussion of these issues, and ideas for moving forward can aid researchers to address these issues in future research. Finally, the samples studied were generally limited (either in terms of lacking in the level of detail reported, or in terms of convenience samples used in university settings). Taking efforts to diversify such samples and to improve reporting detail would be highly informative and could help to better develop moderator analyses in future work.

## Conclusions

Understanding people's personalities is central to many aspects of our lives, from forming and maintaining relationships, to predicting how people's behavior may impact their health, job performance, even political outcomes. As our interactions become increasingly digitized, the way in which we can infer information about people provides opportunities to study human behavior in new and prosperous ways. Digital traces can be used to challenge existing psychological theories with computational methods, and inform data-driven approaches with rich theoretical insights. Together, these advancements have the potential to revolutionize our understanding of human behavior. This two-part study has demonstrated that personality can be predicted from digital traces and that convergent validity is moderate across both human perception and computer-based prediction. Humans were particularly good at identifying Openness and Extraversion online, whereas Neuroticism, Agreeableness, and Conscientiousness appeared to be more challenging. Although convergent validity appeared to be lower for computers across all traits, computers were also good at identifying Openness, but were limited when it came to identifying Neuroticism and Conscientiousness. These insights reinforce the theoretical perspective that internalized traits are more difficult to discern than externalized traits, extending this understanding to the digital realm. Although research in these areas has become increasingly popular in recent years, our synthesis has highlighted that our true understanding of personality (and its relationship to digital traces) is in its infancy (as evidenced by our limited explorations of the predictability of different types of data). As such, we know extremely little about how personality manifests across different types of digital traces beyond social media, and across other forms of digital devices. Future research could exploit these opportunities to develop our understanding of personality across many areas of society.

## References

References marked with an asterisk indicate studies included in the meta-analysis or "hybrid" literature review.

*Adali, S., & Golbeck, J. (2012). *Predicting personality with social behavior* [Conference session]. Advances in Social Networks Analysis and Mining, Istanbul, Turkey. https://doi.org/10.1109/ASONAM.2012.58

*Adalı, S., & Golbeck, J. (2014). Predicting personality with social behavior: A comparative study. *Social Network Analysis and Mining*, *4*(1), Article 159. https://doi.org/10.1007/s13278-014-0159-7

*Adi, G. Y. N. N., Tandio, M. H., Ong, V., & Suhartono, D. (2018). Optimization for automatic personality recognition on Twitter in Bahasa Indonesia. *Procedia Computer Science*, *135*, 473–480. https://doi.org/10.1016/j.procs.2018.08.199

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

*Al Marouf, A., Hasan, M. K., & Mahmud, H. (2019). *Identifying neuroticism from user generated content of social media based on psycholinguistic cues* [Conference session]. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh. https://ieeexplore.ieee.org/abstract/document/8679505

*Al Marouf, A., Hasan, M. K., & Mahmud, H. (2020). Comparative analysis of feature selection algorithms for computational personality prediction from social media. *IEEE Transactions on Computational Social Systems*, *7*(3), 587–599. https://doi.org/10.1109/TCSS.2020.2966910

*Alam, F., & Riccardi, G. (2014). *Predicting personality traits using multimodal information* [Conference session]. ACM Multi Media on Workshop on Computational Personality Recognition, WCPR '14, Orlando, FL, United States. https://dl.acm.org/doi/abs/10.1145/2659522.2659531?casa_token=kHVSURY23i8AAAAA:KQmJb8aGXalIAPP3NbSUBl_ygnM1dUMBLxBFmiS-CsUvAkjlOD7_sXmEVXUlZiZbmljsofOeq_RqHA

*Alam, F., Stepanov, E. A., & Riccardi, G. (2013). *Personality traits recognition on social network—Facebook* [Conference session]. AAAI Conference on Weblogs and Social Media (ICWSM), Cambridge, MA, United States. https://cdn.aaai.org/ocs/6167/6167-30316-1-PB.pdf

Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *55*(3), 387–395. https://doi.org/10.1037/0022-3514.55.3.387

Alharbi, A., Dong, H., Yi, X., Tari, Z., & Khalil, I. (2021). Social media identity deception detection: A survey. *ACM Computing Surveys*, *54*(3), 1–35. https://doi.org/10.1145/3446372

Allport, G. W. (1937). *Personality: A psychological interpretation*. Holt.

*Alsadhan, N., & Skillicorn, D. (2017). *Estimating personality from social media posts* [Poster presentation]. IEEE International Conference on Data Mining Workshops, New Orleans, LA, United States. https://ieeexplore.ieee.org/abstract/document/8215683

*Álvarez-Carmona, M. A., López-Monroy, A. P., Montes-Y-Gómez, M., Villaseñor-Pineda, L., & Escalante, H. J. (2015). *INAOE's participation at PAN'15: Author profiling task* [Conference session]. CEUR Workshop Proceedings, Toulouse, France. https://ceur-ws.org/Vol-1391/122-CR.pdf

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256–274. https://doi.org/10.1037/0033-2909.111.2.256

American Psychological Association. (2010). *APA sixth edition*. https://doi.org/10.1006/mgme.2001.3260

Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, *26*(6), 1289–1295. https://doi.org/10.1016/j.chb.2010.03.018

*Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). "Just the Way You Are": Linking music listening on spotify and personality. *Social Psychological and Personality Science*, *12*(4), 561–572. https://doi.org/10.1177/1948550620923228

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), Article e0139004. https://doi.org/10.1371/journal.pone.0139004

*Appling, D. S., Briscoe, E. J., Hayes, H., & Mappus, R. L. (2013). *Towards automated personality identification using speech acts* [Conference session]. Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, United States. https://ojs.aaai.org/index.php/ICWSM/article/view/14469

*Arnoux, P. H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017). *25 tweets to know you: A new model to predict personality with social media* [Conference session]. Eleventh International AAAI Conference on Web and Social Media (ICWSM), Montreal, QC, Canada. https://ojs.aaai.org/index.php/ICWSM/article/view/14963

*Arroju, M., Hassan, A., & Farnadi, G. (2015). *Age, gender and personality recognition using tweets in a multilingual setting* [Conference session]. Evaluation Forum (CLEF 2015): Experimental IR meets multilinguality, multimodality, and interaction, Toulouse, France. https://ceur-ws.org/Vol-1391/57-CR.pdf

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

*Ates, U. (2014). *Inference of personality using social media profiles*. Middle East Technical University.

*Aydin, B., Kindiroglu, A. A., Aran, O., & Akarun, L. (2016). *Automatic personality prediction from audiovisual data using random forest regression* [Conference session]. Pattern Recognition (ICPR), Cancun, Mexico. https://ieeexplore.ieee.org/abstract/document/7899604

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–158. https://doi.org/10.1016/j.paid.2017.12.018

*Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). *Personality and patterns of Facebook usage* [Conference session]. 4th Annual ACM Web Science Conference, Evanston, IL, United States. https://dl.acm.org/doi/abs/10.1145/2380718.2380722?casa_token=kx6tsKCYGBYAAAAA:T-WnzrynoVVj6WWDmXZp8R9El-luZk78cTthhDm9vtP3Ga0f4NK56WyqInkFOMiVJ_3pqKRDUD6UXg

*Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de? Inferring personality from e-mail addresses. *Journal of Research in Personality*, *42*(4), 1116–1122. https://doi.org/10.1016/j.jrp.2008.02.001

*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. https://doi.org/10.1177/0956797609360756

Bagrow, J. P., Liu, X., & Mitchell, L. (2019). Information flow reveals prediction limits in online social activity. *Nature Human Behaviour*, *3*(2), 122–128. https://doi.org/10.1038/s41562-018-0510-5

*Bai, S., Hao, B., Li, A., Yuan, S., Gao, R., & Zhu, T. (2013). *Predicting big five personality traits of microblog users* [Conference session]. International Conference on Web Intelligence, Atlanta, GA, United States. https://ieeexplore.ieee.org/abstract/document/6690057

*Bai, S., Zhu, T., & Cheng, L. (2012). *Big-five personality prediction based on user behaviors at social network sites*. ArXiv. https://doi.org/10.48550/arXiv.1204.4809

*Baik, J., Lee, K., Lee, S., Kim, Y., & Choi, J. (2016). Predicting personality traits related to consumer behavior using SNS analysis. *New Review of Hypermedia and Multimedia*, *22*(3), 189–206. https://doi.org/10.1080/13614568.2016.1152313

Bakir, V. (2020). Psychological operations in digital political campaigns: Assessing cambridge analytica's psychographic profiling and targeting. *Frontiers in Communication*, *5*, Article 67. https://doi.org/10.3389/fcomm.2020.00067

*Bartoli, A., De Lorenzo, A., Laderchi, A., Medvet, E., & Tarlao, F. (2015). *An author profiling approach based on language-dependent content and stylometric features* [Conference session]. CEUR Workshop Proceedings, Toulouse, France. https://arts.units.it/handle/11368/2846876

*Bayot, R., Gonçalves, T., & Quaresma, P. (2015). *Author profiling of Twitter users* [Conference session]. Notebook for PAN at CLEF 2015, Toulouse, France. https://ceur-ws.org/Vol-1391/123-CR.pdf

Beer, A., & Vazire, S. (2017). Evaluating the predictive validity of personality trait judgments using a naturalistic behavioral criterion: A preliminary test of the self-other knowledge asymmetry model. *Journal of Research in Personality*, *70*, 107–121. https://doi.org/10.1016/j.jrp.2017.06.004

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment*, *90*(3), 250–260. https://doi.org/10.1080/00223890701884970

*Biel, J. I., & Gatica-Perez, D. (2013). The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Transactions on Multimedia*, *15*(1), 41–55. https://doi.org/10.1109/TMM.2012.2225032

*Biel, J. I., & Gatica-Perez, D. (2014). Mining crowdsourced first impressions in online social video. *IEEE Transactions on Multimedia*, *16*(7), 2062–2074. https://doi.org/10.1109/TMM.2014.2346471

*Biel, J.-I., Teijeiro-Mosquera, L., & Gatica-Perez, D. (2012). *FaceTube: Predicting personality from facial expressions of emotion in online conversational video* [Conference session]. 14th ACM International Conference on Multimodal Interaction, Santa Monica, CA, United States. https://doi.org/10.1145/2388676.2388689

*Biel, J.-I., Tsiminaki, V., Dines, J., & Gatica-Perez, D. (2013). *Hi YouTube! Personality impressions and verbal content in social video* [Conference session]. Multimodal Interaction—ICMI '13, Sydney, Australia. https://dl.acm.org/doi/abs/10.1145/2522848.2522877?casa_token=jMMcktNTpGQAAAAA:FyNqkJZeutFTr2wlmwi_cN1Z7Yj3SMgnVk6YqEuIFfklJ-SnFHdHxMohaNNciNEURqMCkKEyTDRXOA

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, *92*(1), 119–135. https://doi.org/10.1037/0022-3514.92.1.119

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*(2), 190–203. https://doi.org/10.1177/1088868318772990

Borenstein, M., Hedges, L. v., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: $I^2$ is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*(4), 645–657. https://doi.org/10.1037/0022-3514.62.4.645

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*(4), 599–614. https://doi.org/10.1037/0022-3514.86.4.599

Bornstein, R. F. (2009). Heisenberg, Kandinsky, and the heteromethod convergence problem: Lessons from within and beyond psychology. *Journal of Personality Assessment*, *91*(1), 1–8. https://doi.org/10.1080/00223890802483235

Bowden-Green, T., Hinds, J., & Joinson, A. (2020). How is extraversion related to social media use? A literature review. *Personality and Individual Differences*, *164*, Article 110040. https://doi.org/10.1016/j.paid.2020.110040

Bowden-Green, T., Hinds, J., & Joinson, A. (2021). Understanding neuroticism and social media: A systematic review. *Personality and Individual Differences*, *168*, Article 110344. https://doi.org/10.1016/j.paid.2020.110344

Brandtzæg, P. B. (2012). Social networking sites: Their users and social implications—A longitudinal study. *Journal of Computer-Mediated*

*Communication*, *17*(4), 467–488. https://doi.org/10.1111/j.1083-6101.2012.01580.x

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press. https://doi.org/10.1525/9780520350519

Brysbaert, M., Bakk, Z., Buchanan, E. M., Drieghe, D., Frey, A., Kim, E., Kuperman, V., Madan, C. R., Marelli, M., Mathôt, S., Valdivia, D. S., & Yap, M. (2021). Into a new decade. *Behavior Research Methods*, *53*, 1–3. https://doi.org/10.3758/s13428-020-01497-y

Buettner, R. (2017). Getting a job via career-oriented social networking markets. *Electronic Markets*, *27*(4), 371–385. https://doi.org/10.1007/s12525-017-0248-3

*Buettner, R. (2017). Predicting user behavior in electronic markets based on personality-mining in large online social networks: A personality-based product recommender framework. *Electronic Markets*, *27*(3), 247–265. https://doi.org/10.1007/s12525-016-0228-z

Buhmann, A., Paßmann, J., & Fieseler, C. (2020). Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse. *Journal of Business Ethics*, *163*, 265–280. https://doi.org/10.1007/s10551-019-04226-4

Butcher, J. N., & Williams, C. J. (1992). *Essentials of MMPI-2 and MMPI-A interpretation*. University of Minnesota Press.

Cadwalladr, C. (2018). Facebook suspends data firm hired by Vote Leave over alleged Cambridge Analytica ties. *The Guardian*. https://www.theguardian.com/us-news/2018/apr/06/facebook-suspends-aggregate-iq-cambridge-analytica-vote-leave-brexit

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, *368*, Article l6890. https://doi.org/10.1136/bmj.l6890

*Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information*, *9*(5), Article 127. https://doi.org/10.3390/info9050127

Carmack, H. J., & Heiss, S. N. (2018). Using the theory of planned behavior to predict college students' intent to use linkedin for job searches and professional networking. *Communication Studies*, *69*(2), 145–160. https://doi.org/10.1080/10510974.2018.1424003

*Celli, F. (2012). *Unsupervised personality recognition for social network sites* [Conference session]. Sixth International Conference on Digital Society, Valencia, Spain. https://www.researchgate.net/profile/Fabio-Celli/publication/258045593_Unsupervised_Personality_Recognition_for_Social_Network_Sites/links/00b7d526b90b4d0923000000/Unsupervised-Personality-Recognition-for-Social-Network-Sites.pdf

*Celli, F., Bruni, E., & Lepri, B. (2014). *Automatic personality and interaction style recognition from Facebook profile pictures* [Conference session]. ACM International Conference on Multimedia—MM '14, Orlando, FL, United States. https://dl.acm.org/doi/abs/10.1145/2647868.2654977?casa_token=NTZ5pxmwhxUAAAAA:0n2osIkAzUg8ltb5itqJKBR2AMZEFnkOHP23R7GBJelA-k8csLxIEqzwvCwclig8SIApCXYiG1tCww

Chan-Olmsted, S. M., Cho, M., & Lee, S. (2013). User perceptions of social media: A comparative study of perceived characteristics and user profiles by social media. *Online Journal of Communication and Media Technologies*, *3*(4). 149–178. https://doi.org/10.29333/ojcmt/2451

*Chapsky, D. (2011). *Leveraging online social networks and external data sources to predict personality* [Conference session]. Advances in Social Networks Analysis and Mining, Kaohsiung, Taiwan. https://ieeexplore.ieee.org/abstract/document/5992637

*Chen, B., Escalera, S., Guyon, I., Ponce-López, V., Shah, N., & Simón, M. O. (2016). Overcoming calibration problems in pattern labeling with pairwise ratings: Application to personality traits. In G. Hua, & H. Jégou (Eds.), *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (pp. 419–432). Springer. https://doi.org/10.1007/978-3-319-49409-8_33

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Chiang, J. J., Lam, P. H., Chen, E., & Miller, G. E. (2022). Psychological stress during childhood and adolescence and its association with inflammation across the lifespan: A critical review and meta-analysis. *Psychological Bulletin*, *148*(1–2), 27–66. https://doi.org/10.1037/bul0000351

*Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, *17*(3), 433–450. https://doi.org/10.1007/s00779-011-0490-1

*Chittaranjan, G., Jan, B., & Gatica-Perez, D. (2011). *Who's who with big-five: Analyzing and classifying personality traits with smartphones* [Conference session]. International Symposium on Wearable Computers, San Fransisco, CA, United States. https://ieeexplore.ieee.org/abstract/document/5959587

Coluccia, A., Pozza, A., Ferretti, F., Carabellese, F., Masti, A., & Gualtieri, G. (2020). Online romance scams: Relational dynamics and psychological characteristics of the victims and scammers. A scoping review. *Clinical Practice and Epidemiology in Mental Health*, *16*(1), 24–35. https://doi.org/10.2174/1745017902016010024

Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology*, *60*(6), 884–894. https://doi.org/10.1037/0022-3514.60.6.884

Connelly, B. S., & Chang, L. (2016). A meta-analytic multitrait multirater separation of substance and style in Social Desirability Scales. *Journal of Personality*, *84*(3), 319–334. https://doi.org/10.1111/jopy.12161

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092–1122. https://doi.org/10.1037/a0021212

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, *3*(3), 292–313. https://doi.org/10.1111/j.1751-9004.2009.00170.x

*Cristani, M., Vinciarelli, A., Segalin, C., & Perina, A. (2013). *Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis* [Conference session]. 21st ACM International Conference on Multimedia, Barcelona, Spain. https://dl.acm.org/doi/abs/10.1145/2502081.2502280?casa_token=xLRLpm2m-J8AAAAA:dtjY5Rauol9Lvc4h1Lxg-oG9y0paYNJrZP4fdd5wQiMXE24YvJxWYEO2Edqt7F6r7kiMiNx10A39TQ

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. (2023). What's in a badge? A computational reproducibility investigation of the open data badge policy in one issue of Psychological Science. *Psychological Science*, *34*(4), 512–522. https://doi.org/10.1177/09567976221140828

*Cucurull, G., Rodríguez, P., Yazici, V. O., Gonfaus, J. M., Roca, F. X., & González, J. (2018). *Deep inference of personality traits by integrating image and word use in social networks*. ArXiv. https://doi.org/10.48550/arXiv.1802.06757

*Cutler, A., & Kulis, B. (2018). *Inferring human traits from facebook statuses* [Conference session]. International Conference on Social Informatics, St Petersburg, Russia. https://doi.org/10.1007/978-3-030-01129-1_11

Dahlke, J. A., & Wiernik, B. M. (2019). *psychmeta*: An R package for psychometric meta-analysis. *Applied Psychological Measurement*, *43*(5), 415–416. https://doi.org/10.1177/0146621618795933

*Darbyshire, D., Kirk, C., Wall, H. J., & Kaye, L. K. (2016). Don't judge a (face)book by its cover: Exploring judgement accuracy of others' personality on Facebook. *Computers in Human Behavior*, *58*, 380–387. https://doi.org/10.1016/j.chb.2016.01.021

*Darliansyah, A., Naeem, M. A., Mirza, F., & Pears, R. (2019). Sentipede: A smart system for sentiment-based personality detection from short texts. *Journal of Universal Computer Science*, *25*(10), 1323–1352. https://doi.org/10.3217/jucs-025-10-1323

Davidson, B. I., Ellis, D., Stachl, C., Taylor, P., & Joinson, A. (2021). *Measurement practices exacerbate the generalizability crisis: Novel digital measures can help*. Psyarxiv. https://doi.org/10.31234/osf.io/8abzy

Davidson, B. I., & Joinson, A. N. (2021). Shape shifting across social media. *Social Media + Society*, *7*(1), 1–11. https://doi.org/10.1177/2056305121990632

*De Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. (2013). *Predicting personality using novel mobile phone-based metrics* [Conference session]. Social Computing, Behavioral-Cultural Modeling and Prediction: 6th International Conference, SBP 2013, Washington DC, United States. https://doi.org/10.1007/978-3-642-37210-0_6

*de Oliveira, R., Karatzoglou, A., Concejero Cerezo, P., Armenta Lopez de Vicuna, A., & Oliver, N. (2011). *Towards a psychographic user model from mobile phone usage* [Conference session]. CHI' 11 Extended Abstratcs on Human Factors in Computing Systems, Vancouver, BC, Canada. https://dl.acm.org/doi/abs/10.1145/1979742.1979920?casa_token=8EI6ZXmjYkoAAAAA:Cd9T4iPVQXoIOJqMJe3GYntEywbVWXbc1rBCC-K-FaT2iD-WR8JOAtBsezFtoZjJO_LfTbLLNDmvEg

DeAndrea, D. C., & Walther, J. B. (2011). Attributions for inconsistencies between online and offline self-presentations. *Communication Research*, *38*(6), 805–825. https://doi.org/10.1177/0093650210385340

Dennis, S., Yim, H., Gerrett, P., Sreekumar, V., & Stone, B. (2019). A system for collecting and analyzing experience-sampling data. *Behavior Research Methods2*, *51*(4), 1824–1838. https://doi.org/10.3758/s13428-019-01260-y

*Dhall, A., & Hoey, J. (2016). *First impressions—Predicting user personality from twitter profile images* [Conference session]. Human Behavior Understanding: 7th International Workshop, HBU 2016, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-319-46843-3_10

Digman, J. M. (1990). Personality structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, *41*(1), 417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, *13*(2), Article e25025. https://doi.org/10.5964/spb.v13i2.25025

Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences*, *160*, 109956. https://doi.org/10.1016/j.paid.2020.109956

*Dos Santos, V. G., Paraboni, I., & Silva, B. B. C. (2017). *Big five personality recognition from multiple text genres* [Conference session]. Text, Speech, and Dialogue: 20th International Conference, TSD 2017, Prague, Czech Republic. https://doi.org/10.1007/978-3-319-64206-2_4

*Dotti, D., Popa, M., & Asteriadis, S. (2018). *Behavior and personality analysis in a nonsocial context dataset* [Conference session]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, United States. https://openaccess.thecvf.com/content_cvpr_2018_workshops/w48/html/Dotti_Behavior_and_Personality_CVPR_2018_paper.html

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ellis, D. A. (2020). *Smartphones within psychological science*. Cambridge University Press. https://doi.org/10.1017/9781108671408

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, *11*(2), 415–441. https://doi.org/10.1111/j.1083-6101.2006.00020.x

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

*Ergu, I., Zerrin, I., & Yankayis, I. (2019). *Predicting personality with Twitter Data and Machine Learning Models* [Conference session]. 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey. https://ieeexplore.ieee.org/abstract/document/8946355

*Escalante, H. J., Guyon, I., Escalera, S., Jacques, J., Madadi, M., Baro, X., Ayache, S., Viegas, E., Gucluturk, Y., Guclu, U., Van Gerven, M. A. J., & Van Lier, R. (2017). *Design of an explainable machine learning challenge for video interviews* [Conference session]. International Joint Conference on Neural Networks, Archorage, AK, United States. https://ieeexplore.ieee.org/abstract/document/7966320

*Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., Davalos, S., Moens, M. F., & De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, *26*(2–3), 109–142. https://doi.org/10.1007/s11257-016-9171-0

*Farnadi, G., Sushmita, S., Sitaraman, G., Ton, N., De Cock, M., & Davalos, S. (2014). *A multivariate regression approach to personality impression recognition of vloggers* [Conference session]. 2014 ACM Multi Media on Workshop on Computational Personality Recognition—WCPR '14, Orlando, FL, United States. https://dl.acm.org/doi/abs/10.1145/2659522.2659526?casa_token=I1wkCCompEkAAAAA:5pIgCg7szSK_xqK0d24cBIyHgYbBRhDsuUNiTReZ9xvyZWIbL6Y56EtGwNvjktYEuIBUHPD_AizKNQ

*Farnadi, G., Zoghbi, S., Moens, M., & De Cock, M. (2013). *Recognising personality traits using Facebook status updates* [Conference session]. Weblogs and Social Media (ICWSM13), Cambridge, MA, United States. https://ojs.aaai.org/index.php/ICWSM/article/view/14470

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & van den Noortgate, W. (2021). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *Journal of Experimental Education*, *89*(1), 125–144. https://doi.org/10.1080/00220973.2019.1582470

*Ferwerda, B., Schedl, M., & Tkalcic, M. (2016). *Using instagram picture features to predict users' personality* [Conference session]. MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, United States. https://doi.org/10.1007/978-3-319-27671-7_71

*Ferwerda, B., & Tkalcic, M. (2018). *Predicting users' personality from instagram pictures: Using visual and/or content features?* [Conference session]. 26th Conference on User Modeling, Adaptation and Personalization, Singapore. https://dl.acm.org/doi/abs/10.1145/3209219.3209248?casa_token=68WBB90TsX4AAAAA:5CIB-CFltU9QpR0tMw3BngUIzcZV6tfAmSZJWSOva9T5XvcvzK0lwvuRzR2VVESsqbw0OvajHPwsZw

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Florenthal, B. (2015). Applying uses and gratifications theory to students' LinkedIn usage. *Young Consumers*, *16*(1), 17–35. https://doi.org/10.1108/YC-12-2013-00416

Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA*, *105*(2), Article 203. https://doi.org/10.5195/jmla.2017.88

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670. https://doi.org/10.1037/0033-295X.102.4.652

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. Academic Press.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*(3), 177–182. https://doi.org/10.1177/0963721412445309

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, *55*(1), 149–158. https://doi.org/10.1037/0022-3514.55.1.149

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

Funder, D. C., & West, S. G. (1993). Viewpoints on personality: Consensus, self-other agreement, and accuracy in personality judgment. *Journal of Personality*, *61*(4), 457–476. https://doi.org/10.1111/j.1467-6494.1993.tb00778.x

Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: Examination of a functional hypothesis. *Journal of Personality and Social Psychology*, *62*(4), 688–698. https://doi.org/10.1037/0022-3514.62.4.688

*Gao, N., Shao, W., & Flora, D. S. (2019). Predicting personality traits from physical activity intensity. *Computer*, *52*(7), 47–56. https://doi.org/10.1109/MC.2019.2913751

*Gao, R., Hao, B., Bai, S., Li, L., Li, A., & Zhu, T. (2013). *Improving user profile with personality traits predicted from social media content* [Conference session]. Recommender systems—RecSys '13, Hong Kong, China. https://dl.acm.org/doi/abs/10.1145/2507157.2507219?casa_token=Jxkufnbf62MAAAAA:HUaFIdyYYVW-gkeXcWXvFqn5uWzb9E2XS0bcGWVZScrUqAS8D-HB4V763xR6QvttgjSWfhRorr-tZA

Garcia, D. (2017). Leaking privacy and shadow profiles in online social networks. *Science Advances*, *3*(8), Article e1701172. https://doi.org/10.1126/sciadv.1701172

Gerlach, M., Farb, B., Revelle, W., & Nunes Amaral, L. A. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour*, *2*, 735–742. https://doi.org/10.1038/s41562-018-0419-z

Geyer, K., Ellis, D. A., & Piwek, L. (2019). A simple location-tracking app for psychological research. *Behavior Research Methods*, *51*(6), 2840–2846. https://doi.org/10.3758/s13428-018-1164-y

*Ghavami, S. M., Asadpour, M., Hatami, J., & Mahdavi, M. (2015). *Facebook user's like behavior can reveal personality* [Conference session]. Information and Knowledge Technology, IKT 2015, Urmia, Iran. https://ieeexplore.ieee.org/abstract/document/7288797

*Gievska, S., & Koroveshovski, K. (2014). *The impact of affective verbal content on predicting personality impressions in youtube videos* [Conference session]. 2014 ACM Multi Media on Workshop on Computational Personality Recognition, Orlando, FL, United States. https://dl.acm.org/doi/abs/10.1145/2659522.2659529?casa_token=JMSTdq1Y8a4AAAAA:3G2eqmnUXe0xO-OiYFKRafBJNcz5E7gKmqRZ4FIqHBEEaOK_Ndzc8rjVnB-xce9CQMMbNGEuwVCXRA

*Gill, A. J., Nowson, S., & Oberlander, J. (2009). *What are they blogging about? Personality, topic and motivation in blogs* [Conference session]. International AAAI Conference on Weblogs Ans Social Media, San Jose, CA, United States. https://ojs.aaai.org/index.php/ICWSM/article/view/13949

Gilpin, A. R. (1993). Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement*, *53*(1), 87–92. https://doi.org/10.1177/0013164493053001007

*Giménez, M., Hernández, D. I., & Pla, F. (2015). *Segmenting target audiences: Automatic author profiling using tweets* [Conference session]. CEUR Workshop Proceedings, Toulouse, France. https://downloads.webis.de/pan/publications/papers/gimenez_2015.pdf

*Gladstone, J. J., Matz, S. C., & Lemaire, A. (2019). Can psychological traits be inferred from spending? Evidence from transaction data. *Psychological Science*, *30*(7), 1087–1096. https://doi.org/10.1177/0956797619849435

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology*, *56*(11), 1129–1135. https://doi.org/10.1016/S0895-4356(03)00177-X

*Golbeck, J. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research*, *2*, Article 2. https://doi.org/10.17705/1atrr.00009

*Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). *Predicting personality from twitter* [Conference session]. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. Boston, MA, United States. https://ieeexplore.ieee.org/abstract/document/6113107

*Golbeck, J., Robles, C., & Turner, K. (2011). *Predicting personality with social media* [Conference session]. Annual Conference Extended Abstracts on Human Factors in Computing Systems—CHI EA '11. Vancouver, BC, Canada. https://dl.acm.org/doi/abs/10.1145/1979742.1979614?casa_token=sHNXcPQl9HUAAAAA:mRvVMzGhyXoX1NWODJ2TJUCuDKSfSZSnZNHhibP3CcWzd0eRYM9m-qIKlN-_aMFUlL0jnXcPVQCe0A

*Gonzalez-Gallardo, C. E., Montes, A., Sierra, G., Nunez-Juarez, J. A., Salinas-Lopez, A. J., & Ek, J. (2015). *Tweets classification using corpus dependent tags, character and POS N-grams* [Conference session]. CEUR Workshop Proceedings, Toulouse, France. https://ceur-ws.org/Vol-1391/104-CR.pdf

*Gorbova, J., Lusi, I., Litvin, A., & Anbarjafari, G. (2017). *Automated screening of job candidate based on multimodal video processing* [Conference session]. Computer Vision and Pattern Recognition Workshops, Honolulu, HI, United States. https://openaccess.thecvf.com/content_cvpr_2017_workshops/w26/html/Gorbova_Automated_Screening_of_CVPR_2017_paper.html

*Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, *14*(9), 483–488. https://doi.org/10.1089/cyber.2010.0087

*Gosling, S. D., Gaddis, S., & Vazire, S. (2007). *Personality impressions based on Facebook profiles* [Conference session]. International Conference on Weblogs and Social Media (ICWSM), Boulder, CO, United States. https://icwsm.org/papers/3--Gosling-Gaddis-Vazire.pdf

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379–398. https://doi.org/10.1037/0022-3514.82.3.379

*Gou, L., Zhou, M. X., & Yang, H. (2014). *KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences* [Conference session]. SIGCHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/2556288.2557398

Gough, H. G. (2012). The California psychological inventory. In C. E. Watkins, Jr. & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (2nd ed., pp. 45–71). Lawrence Erlbaum.

Gough, H. G., & Bradley, P. (1996). *CPI manual* (3rd ed.). Consulting Psychologists Press.

*Graham, L. T., & Gosling, S. D. (2012). Impressions of world of Warcraft players' personalities based on their usernames: Interobserver consensus but no accuracy. *Journal of Research in Personality*, *46*(5), 599–603. https://doi.org/10.1016/j.jrp.2012.05.002

*Grover, T., & Mark, G. (2017). *Digital footprints: Predicting personality from temporal patterns of technology use* [Conference session]. ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable

Computers, Maui, HI, United States. https://dl.acm.org/doi/abs/10.1145/3123024.3123139

*Guan, Z., Wu, B., Wang, B., & Liu, H. (2020). *Personality2vec: Network representation learning for personality* [Conference session]. 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). https://ieeexplore.ieee.org/abstract/document/9172426

*Guan, Z., Wu, B., Wang, B., & Liu, H. (2020). *Personality2vec: Network representation learning for personality* [Conference session]. 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), Zhuhai, China. https://ieeexplore.ieee.org/abstract/document/9172426

*Güçlütürk, Y., Güçlü, U., Baró, X., Escalante, H. J., Guyon, I., Escalera, S., van Gerven, M. A. J., & Van Lier, R. (2018). Multimodal first impression analysis with deep residual networks. *IEEE Transactions on Affective Computing*, *9*(3), 316–329. https://doi.org/10.1109/TAFFC.2017.2751469

*Güçlütürk, Y., Güçlü, U., van Gerven, M. A. J., & van Lier, R. (2016). *Deep impression: Audiovisual deep residual networks for multimodal apparent personality trait recognition* [Conference session]. Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8 10 and 15–16, 2016, Proceedings, Part III 14, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-319-49409-8_28

*Guntuku, S. C., Qiu, L., Roy, S., Lin, W., & Jakhetiya, V. (2015). *Do others perceive you as you want them to? Modeling personality based on selfies* [Conference session]. 1st International Workshop on Affect & Sentiment in Multimedia, Brisbane, Australia. https://doi.org/10.1145/2813524.2813528

*Guntuku, S. C., Zhou, J. T., Roy, S., Lin, W., & Tsang, I. W. (2018). Who likes what and, why? Insights into modeling users' personality based on image likes. *IEEE Transactions on Affective Computing2*, *9*(1), 130–143. https://doi.org/10.1109/TAFFC.2016.2581168

*Gürpinar, F., Kaya, H., & Salah, A. A. (2017). *Multimodal fusion of audio, scene, and face features for first impression estimation* [Conference session]. Pattern Recognition (ICPR), Cancun, Mexico. https://doi.org/10.1109/ICPR.2016.7899605

*Halim, Z., Atif, M., Rashid, A., & Edwin, C. A. (2017). profiling players using real-world datasets: Clustering the data and correlating the results with the big-five personality traits. *IEEE Transactions on Affective Computing*, *10*(4), 568–584. https://doi.org/10.1109/TAFFC.2017.2751602

*Hall, A. N., & Matz, S. C. (2020). Targeting item-level nuances leads to small but robust improvements in personality prediction from digital footprints. *European Journal of Personality*, *34*(5), 873–884. https://doi.org/10.1002/per.2253

*Hall, J. A., Pennington, N., & Lueders, A. (2014). Impression management and formation on Facebook: A lens model approach. *New Media & Society*, *16*(6), 958–982. https://doi.org/10.1177/1461444813495166

Harari, G. M., Vaid, S. S., Müller, S. R., Stachl, C., Marrero, Z., Schoedel, R., Bühner, M., & Gosling, S. D. (2020). Personality sensing for theory development and assessment in the digital age. *European Journal of Personality*, *34*(5), 649–669. https://doi.org/10.1002/per.2273

Harrer, M., Cuijpers, P., Furukawa, A., & Ebert, D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

*Hassanein, M., Hussein, W., Rady, S., & Gharib, T. F. (2018). *Predicting personality traits from Social media using text semantics* [Conference session]. Computer Engineering and Systems (ICCES), Cairo, Egypt. https://doi.org/10.1109/ICCES.2018.8639408

Hathaway, S. R., & McKinley, J. C. (1943). *Manual for the Minnesota Multiphasic Personality Inventory*. Psychological Corporation.

Heller, D., Watson, D., & Hies, R. (2004). The role of person versus situation in life satisfaction: A critical examination. *Psychological Bulletin*, *130*(4), 574–600. https://doi.org/10.1037/0033-2909.130.4.574

Hendrikx, F., Bubendorfer, K., & Chard, R. (2015). Reputation systems: A survey and taxonomy. *Journal of Parallel and Distributed Computing*, *75*, 184–197. https://doi.org/10.1016/j.jpdc.2014.08.004

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). Most people are not WEIRD. *Nature*, *466*, Article 29. https://doi.org/10.1038/466029a

*Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2022). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*, *107*(8), 1323–1351. https://doi.org/10.1037/apl0000695

*Hickman, L., Tay, L., & Woo, S. E. (2019). Validity evidence for off-the-shelf language-based personality assessment using video interviews: Convergent and discriminant relationships with self and observer ratings. *Personnel Assessment and Decisions*, *5*(3), Article 3. https://doi.org/10.25035/pad.2019.03.003

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, *37*(5), 1158–1160. https://doi.org/10.1093/ije/dyn204

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hinds, J., & Joinson, A. (2019). Human and computer personality prediction from digital footprints. *Current Directions in Psychological Science*, *28*(2), 204–211. https://doi.org/10.1177/0963721419827849

Hinds, J. & Joinson, A. N. (2024, January 3). *Personality meta-analysis*. https://osf.io/9cndh/

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, *143*, Article 102498. https://doi.org/10.1016/j.ijhcs.2020.102498

Hofmann, W., & Patel, P. (2015). SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. *Social Science Computer Review*, *33*(2), 235–253. https://doi.org/10.1177/0894439314525117

Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, *30*(6), 377–386. https://doi.org/10.1177/0270467610385893

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

*Howlader, P., Pal, K. K., Cuzzocrea, A., & Kumar, S. D. M. (2018). *Predicting facebook-users' personality based on status and linguistic features via flexible regression analysis techniques* [Conference session]. 33rd Annual ACM Symposium on Applied Computing, Pau, France. https://doi.org/10.1145/3167132.3167166

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge. https://doi.org/10.4324/9781315650982

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, *8*(4), 275–292. https://doi.org/10.1111/1468-2389.00156

*Iacobelli, F., & Culotta, A. (2013). *Too neurotic, not too friendly: Structured personality classification on textual data* [Conference session]. Seventh International AAAI Conference on Weblogs and Social Media (ICWSM), Cambidge, MA, United States.

*Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). *Large scale personality classification of bloggers* [Conference session]. Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, United States. https://doi.org/10.1007/978-3-642-24571-8_71

*Imran, A., Faiyaz, M., & Akhtar, F. (2018). An enhanced approach for quantitative prediction of personality in Facebook posts. *International*

*Journal of Education and Management Engineering*, *8*(2), 8–19. https://doi.org/10.5815/ijeme.2018.02.02

*Indira, R., & Maharani, W. (2021). *Personality detection on social media Twitter using long short-term memory with Word2Vec* [Conference session]. Communication, Networks and Satellite, Comnetsat, Purwokerto, Indonesia. https://doi.org/10.1109/COMNETSAT53002.2021.9530820

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7), Article e010247. https://doi.org/10.1136/bmjopen-2015-010247

Ioannidis, J. P. A. (2005). Differentiating biases from genuine heterogeneity: Distinguishing artifactual from substantive effects. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 287–302). Wiley. https://doi.org/10.1002/0470870168.ch15

Ioannidis, J. P. A., Cappelleri, J. C., & Lau, J. (1998). Issues in comparisons between meta-analyses and large trials. *JAMA*, *279*(14), 1089–1093. https://doi.org/10.1001/jama.279.14.1089

*Iqbal, H. R., Ashraf, M. A., & Nawab, R. M. A. (2015). *Predicting an author's demographics from text using topic modeling approach* [Conference session]. CLEF 2015 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings, Toulouse, France.

*Ishikawa, L. T. (2009). *Lay assessments of personality traits and depression in MySpace* [Doctoral dissertation]. The University of Arizona.

Jackson, N., & Waters, E. (2005). Criteria for the systematic review of health promotion and public health interventions. *Health Promotion International*, *20*(4), 367–374. https://doi.org/10.1093/heapro/dai022

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12. https://doi.org/10.1016/0197-2456(95)00134-4

*Jarvis, M. M. (2010). *Facebook and personality: What do status updates really communicate?* [Bachelor's thesis, University of Arizona]

Jensen, C., Davis, J., & Farnham, S. (2002). *Finding others online: Reputation systems for social online spaces* [Conference session]. Human Factors in Computing Systems, Minneapolis, MN, United States.

*Jeremy, N. H., Prasetyo, C., & Suhartono, D. (2019). Identifying personality traits for Indonesian user from twitter dataset. *International Journal of Fuzzy Logic and Intelligent Systems*, *19*(4), 283–289. https://doi.org/10.5391/IJFIS.2019.19.4.283

Jin, S. A. A., & Phua, J. (2014). Following celebrities' tweets about brands: The impact of Twitter-based electronic word-of-mouth on consumers source credibility perception, buying intention, and social identification with celebrities. *Journal of Advertising*, *43*(2), 181–195. https://doi.org/10.1080/00913367.2013.827606

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*(4), 521–551. https://doi.org/10.1111/j.1467-6494.1993.tb00781.x

Johnson, G. D. (2021). Joseph Henrich, The WEIRDest people in the world: How the west became psychologically peculiar and particularly prosperous. *International Sociology*, *36*(5), 780–782. https://doi.org/10.1177/02685809211057670

Joshi, M., Dredze, M., Cohen, W., & Rose, C. (2012). *Multi-domain learning: When do domains matter?* [Conference session]. Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea.

*Junior, R. A., & Inkpen, D. (2017). *Using cognitive computing to get insights on personality traits from twitter messages* [Conference session]. Canadian Conference on Artificial Intelligence, Edmonton, Canada. https://doi.org/10.1007/978-3-319-57351-9_32

*Kafeza, E., Kanavos, A., Makris, C., & Vikatos, P. (2014). *T-PICE: Twitter personality based influential communities extraction system* [Conference session]. 2014 IEEE International Congress on Big Data, BigData Congress 2014, Washington, DC, United States. https://doi.org/10.1109/BigData.Congress.2014.38

*Kambham, N. K., Stanley, K. G., & Bell, S. (2018). *Predicting personality traits using smartphone sensor data and app usage data* [Conference session]. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada. https://doi.org/10.1109/IEMCON.2018.8614854

*Kamisaka, D., & Ishikawa, Y. (2021). *Personality prediction with cross-modality feature projection* [Conference session]. 2021 International Conference on Multimodal Interaction (ICMI), Montreal, Canada. https://doi.org/10.1145/3462244.3479948

*Kampman, O., Barezi, E. J., Bertero, D., & Fung, P. (2018). *Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction* [Conference session]. 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), Melbourne, Australia. https://doi.org/10.18653/v1/P18-2096

*Karanatsiou, D., Sermpezis, P., Gruda, D., Kafetsios, K., Dimitriadis, I., & Vakali, A. (2022). My tweets bring all the traits to the yard: Predicting personality and relational traits in online social networks. *ACM Transactions on the Web*, *16*(2), 1–26. https://doi.org/10.1145/3523749

Kedar, S. v., & Bormane, D. S. (2015). *Automatic personality assessment: A systematic review* [Conference session]. 2015 International Conference on Information Processing (ICIP), Pune, India. https://doi.org/10.1109/INFOP.2015.7489402

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, *98*(2), 155–163. https://doi.org/10.1037/0033-295X.98.2.155

Kenny, D. A. (1994). *Interpersonal percdeption: A social relations analysis*. Guilford Press.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, *8*(3), 265–280. https://doi.org/10.1207/s15327957pspr0803_3

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), 1123–1134. https://doi.org/10.1037/0003-066X.39.10.1123

*Kim, Y., & Kim, J. H. (2018). Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Information Processing & Management*, *54*(6), 1101–1114. https://doi.org/10.1016/j.ipm.2018.07.005

*Kiprov, Y., Hardalov, M., Nakov, P., & Koychev, I. (2015). *SU @ PAN 2015<r_space/>: Experiments in author profiling* [Conference session]. CLEF 2015 Labs and Workshops. Notebook Papers, Toulouse, France.

Kitchgaessner, S. (2018). Cambridge Analytica used data from Facebook and Politico to help Trump. *The Guardian*. https://www.theguardian.com/technology/2017/oct/26/cambridge-analytica-used-data-from-facebook-and-politico-to-help-trump

*Kleanthous, S., Herodotou, C., Samaras, G., & Panayiotis, G. (2016). *Detecting personality traces in users' social activity* [Conference session]. Social Computing and Social Media, Toronto, ON, Canada.

*Kluemper, D. H., Rosen, P. A., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology*, *42*(5), 1143–1172. https://doi.org/10.1111/j.1559-1816.2011.00881.x

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

*Kocher, M. (2015). *Unine at CLEF 2015: Author profiling—Notebook for PAN at CLEF 2015* [Conference session]. CLEF 2015 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings, Toulouse, France.

*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. https://doi.org/10.1007/s10994-013-5415-y

*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 of the United States of America*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Krämer, N. C., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media Psychology*, *20*(3), 106–116. https://doi.org/10.1027/1864-1105.20.3.106

Krause, A. E., North, A. C., & Heritage, B. (2014). The uses and gratifications of using Facebook music listening applications. *Computers in Human Behavior*, *39*, 71–77. https://doi.org/10.1016/j.chb.2014.07.001

*Kumar, K. N. P., & Gavrilova, M. L. (2022). Latent personality traits assessment from social network activity using contextual language embedding. *IEEE Transactions on Computational Social Systems*, *9*(2), 638–649. https://ieeexplore.ieee.org/abstract/document/9531972

*Kumar, U., Reganti, A. N., Maheshwari, T., Chakroborty, T., Gambäck, B., & Das, A. (2018). Inducing personalities and values from language use in social network communities. *Information Systems Frontiers*, *20*(6), 1219–1240. https://doi.org/10.1007/s10796-017-9793-8

*Kunte, A. V., & Panicker, S. (2019). *Using textual data for personality prediction: A machine learning approach* [Conference session]. Information Systems and Computer Networks (ISCON), Muthura, India. https://ieeexplore.ieee.org/abstract/document/9036220

Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self-other agreement on personality ratings in college roommates. *Journal of Personality*, *71*(1), 21–48. https://doi.org/10.1111/1467-6494.t01-1-00005

*Küster, L., Trahms, C., & Voigt-Antons, J. N. (2018). *Predicting personality traits from touchscreeen based interactions* [Conference session]. Quality of Multimedia Experience (QoMEX), Cagliari, Italy. https://ieeexplore.ieee.org/abstract/document/8463375

*Lai, S., Sun, B., Wu, F., & Xiao, R. (2020). Automatic personality identification using students' online learning behavior. *IEEE Transactions on Learning Technologies*, *13*(1), 26–37. https://doi.org/10.1109/TLT.2019.2924223

*Laleh, A., & Shahram, R. (2017). *Analyzing Facebook activities for personality recognition* [Conference session]. Machine Learning and Applications (ICMLA), Cancun, Mexico, Cancun, Mexico. https://doi.org/10.1109/ICMLA.2017.00-29

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. *Flow and the foundations of positive psychology* (pp. 21–34). Springer. https://doi.org/10.1007/978-94-017-9088-8_2

Larsson, S., Jensen-Urstad, A., & Heintz, F. (2021). Notified but unaware: Third party tracking online. *Critical Analysis of Law*, *8*(1). https://doi.org/10.33137/cal.v8i1.36282

Lee, E., Ahn, J., & Kim, Y. J. (2014). Personality traits and self-presentation at Facebook. *Personality and Individual Differences*, *69*, 162–167. https://doi.org/10.1016/j.paid.2014.05.020

*Lepri, B., Staiano, J., Shmueli, E., Pianesi, F., & Pentland, A. (2016). The role of personality in shaping social networks and mediating behavioral change. *User Modeling and User-Adapted Interaction*, *26*(2–3), 143–175. https://doi.org/10.1007/s11257-016-9173-y

*Li, C., Wan, J., & Wang, B. (2017). *Personality prediction of social network users* [Conference session]. Distributed Computing and Applications to Business, Engineering and Science (DCABES), Anyang, China. https://doi.org/10.1109/DCABES.2017.25

*Li, J., & Chignell, M. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human Computer Studies*, *68*(9), 589–602. https://doi.org/10.1016/j.ijhcs.2010.04.001

*Li, L., Li, A., Hao, B., Guan, Z., & Zhu, T. (2014). Predicting active users' personality based on micro-blogging behaviors. *PLOS ONE*, *9*(1), Article e84997. https://doi.org/10.1371/journal.pone.0084997

Li, S., Feng, B., Li, N., & Tan, X. (2015). How social context cues in online support-seeking influence self-disclosure in support provision. *Communication Quarterly*, *63*(5), 586–602. https://doi.org/10.1080/01463373.2015.1078389

*Lin, J., Mao, W., & Zeng, D. D. (2017). Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems*, *132*, 204–214. https://doi.org/10.1016/j.knosys.2017.06.031

Lippa, R. (1998). The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis. *Journal of Research in Personality*, *32*(1), 80–107. https://doi.org/10.1006/jrpe.1997.2189

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. In L. Blackman & D. J. Rog (Eds.), *Applied social research methods series* (Vol. 49, pp. 1–264). Sage Publications.

*Liu, L., Preot, D., & Ungar, L. (2016). *Analyzing personality through social media profile picture choice*. Association for the Advancement of Artificial Intelligence. https://doi.org/10.4103/0019-5545.166636

*Liu, X., Nie, D., Bai, S., Hao, B., & Zhu, T. (2014). *Personality prediction for microblog users with active learning method* [Conference session]. International Conference on Human Centered Computing, Phnonm Penh, Cambodia.

*Liu, X., & Zhu, T. (2016). Deep learning for constructing microblog behavior representation to identify social media user's personality. *PeerJ Computer Science*, *2*, Article e81. https://doi.org/10.7717/peerj-cs.81

*Liu, Y., Wang, J., & Jiang, Y. (2016). PT-LDA: A latent variable model to predict personality traits of social network users. *Neurocomputing*, *210*, 155–163. https://doi.org/10.1016/j.neucom.2015.10.144

Loevinger, J., Gleser, G. C., & DuBois, P. H. (1953). Maximizing the discriminating power of a multiple-score test. *Psychometrika*, *18*(4), 309–317. https://doi.org/10.1007/BF02289266

Loiacono, E. T. (2015). Self-disclosure behavior on social networking web sites. *International Journal of Electronic Commerce*, *19*(2), 66–94. https://doi.org/10.1080/10864415.2015.979479

*Lucky, H., & Suhartono, D. (2021). *Towards classification of personality prediction model: A combination of BERT word embedding and MLSMOTE* [Conference session]. Computer Science and Artificial Intelligence (ICCSAI), Jakarta, Indonesia. https://doi.org/10.1109/ICCSAI53272.2021.9609750

Luft, J., & Ingham, H. (1955). *The Johari window, a graphic model of interpersonal awareness* [Conference session]. Proceedings of the Western Training Laboratory in Group Development, University of California, Los Angeles, CA, United States.

*Maharjan, S., & Solorio, T. (2015). *Using wide range of features for author profiling* [Conference session]. Working Notes Papers of the CLEF 2015 Evaluation Labs, CEUR Workshop Proceedings, Toulouse, France. https://ceur-ws.org/Vol-1391/153-CR.pdf

*Mao, Y., Zhang, D., Wu, C., Zheng, K., & Wang, X. (2018). *Feature analysis and optimisation for computational personality recognition* [Conference session]. International Conference on Computer and Communications, Chengdu, China. https://ieeexplore.ieee.org/abstract/document/8780801

*Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal Web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, *90*(6), 1014–1031. https://doi.org/10.1037/0022-3514.90.6.1014

Marengo, D., & Montag, C. (2020). Digital phenotyping of big five personality via Facebook data mining: A meta-analysis. *Digital Psychology*, *1*(1), 52–64. https://doi.org/10.24989/dp.v1i1.1823

Marengo, D., & Settanni, M. (2019). Mining Facebook data for personality prediction: An overview. In H. Baumeister, & C. Montag (Eds.), *Digital phenotyping and mobile sensing. Studies in neuroscience, psychology and behavioral economics* (pp. 121–136). Springer, Cham. https://doi.org/10.1007/978-3-030-31620-4_7

*Markey, P. M., & Wells, S. M. (2002). Interpersonal perception in internet chat rooms. *Journal of Research in Personality*, *36*(2), 134–146. https://doi.org/10.1006/jrpe.2002.2340

*Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013). *Mining Facebook data for predictive personality modeling* [Conference session]. Weblogs and Social Media, Cambridge, MA, United States. https://doi.org/10.1177/0191453711410031

Matz, S. C., & Harari, G. M. (2021). Personality-place transactions: Mapping the relationships between big five personality traits, states, and daily places. *Journal of Personality and Social Psychology*, *120*(5), 1367–1385. https://doi.org/10.1037/pspp0000297

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

*McCollister, C., Huang, S., & Luo, B. (2015). *Building topic models to predict author attributes from Twitter messages* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review*, *19*(2), 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the Five-Factor Model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T., Jr. (1999). A Five-Factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). Guilford Press. https://doi.org/10.1111/j.1600-0706.2011.19910.x

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*, Article e16800. https://doi.org/10.7554/eLife.16800

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Michikyan, M., Dennis, J., & Subrahmanyam, K. (2015). Can you guess who I am? Real, ideal, and false self-presentation on Facebook among emerging adults. *Emerging Adulthood*, *3*(1), 55–64. https://doi.org/10.1177/2167696814532442

Michikyan, M., Subrahmanyam, K., & Dennis, J. (2014). Can you tell who I am? Neuroticism, extraversion, and online self-presentation among young adults. *Computers in Human Behavior*, *33*, 179–183. https://doi.org/10.1016/j.chb.2014.01.010

Mischel, W. (1968). *Personality and assessment*. Wiley.

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189

*Mønsted, B., Mollgaard, A., & Mathiesen, J. (2018). Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality*, *74*, 16–22. https://doi.org/10.1016/j.jrp.2017.12.004

Montag, C., & Elhai, J. D. (2019). A new agenda for personality psychology in the digital age? *Personality and Individual Differences*, *147*, 128–134. https://doi.org/10.1016/j.paid.2019.03.045

*Moraes, R., Pinto, L. L., Pilankar, M., & Rane, P. (2020). *Personality assessment using social media for hiring candidates* [Conference session]. Communication Systems, Computing and IT Applications, Mumbai, India. https://doi.org/10.1109/CSCITA47329.2020.9137818

*Moreno, D. R., Gomez, J. C., Almanza-Ojeda, D., & Ibarra-Manzano, M. (2019). *Prediction of personality traits in Twitter users with latent features* [Conference session]. Electronics, Communications and Computers (CONIELECOMP), Cholulua, Mexico.

*Moreno-Armendáriz, M. A., Martínez, C. A. D., Calvo, H., & Moreno-Sotelo, M. (2020). Estimation of personality traits from portrait pictures using the five-factor model. *IEEE Access*, *8*, 201649–201665. https://ieeexplore.ieee.org/abstract/document/9244051

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*(3), 474–490. https://doi.org/10.1037/pspp0000100

*Mukta, M. S. H., Ali, M. E., & Mahmud, J. (2016). Identifying and validating personality traits-based homophilies for an egocentric network. *Social Network Analysis and Mining*, *6*(1), Article 74. https://doi.org/10.1007/s13278-016-0383-4

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. https://doi.org/10.1002/per.2262

Murray, H. G. (1975). Predicting student ratings of college teaching from peer ratings of personality types. *Teaching of Psychology*, *2*(2), 66–69. https://doi.org/10.1207/s15328023top0202_4

*Najib, F., Cheema, W. A., & Nawab, R. M. A. (2015). *Author's traits prediction on Twitter data using content based approach* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, *35*(12), 1661–1671. https://doi.org/10.1177/0146167209346309

*Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. *Psychological Science*, *29*(7), 1145–1158. https://doi.org/10.1177/0956797618761659

*Ngatirin, N. R., Zainol, Z., & Yoong, T. L. C. (2016). *A comparative study of different classifiers for automatic personality prediction* [Conference session]. Control System, Computing and Engineering (ICCSCE), Penang, Malaysia. https://doi.org/10.1109/ICCSCE.2016.7893613

*Nguyen, T., Phung, D., Adams, B., & Venkatesh, S. (2011). *Towards discovery of influence and personality traits through social link prediction* [Poster presentation]. Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain. https://doi.org/10.1097/01.TP.0000131167.21930.B8

*Nie, D., Guan, Z., Hao, B., Bai, S., & Zhu, T. (2014). *Predicting personality on social media with semi-supervised learning* [Conference session]. IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Warsaw, Poland. https://doi.org/10.1109/WI-IAT.2014.93

*Nie, D., Li, L., & Zhu, T. (2013). *Conscientiousness measurement from Weibo's public information* [Conference session]. Partially Supervised Learning: Second IAPR International Workshop, PSL 2013, Nanjing, China. https://doi.org/10.1007/978-3-642-40705-5_6

*Nie, Y. J., Gao, G. J., Wang, Y. X., Liu, D. X., & Gao, K. (2017). *Personality predicting model based on user's linguistic behavior* [Conference session]. Modelling, Identification and Control (ICMIC), Kunming, China.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, *4*(6), 681–691. https://doi.org/10.1037/h0024002

*Nowson, S., & Gill, A. J. (2014). *Look! who's talking? Projection of extraversion across different social contexts* [Conference session]. 2014 ACM Multi Media on Workshop on Computational Personality Recognition, Orlando, FL, United States.

*Nowson, S., Perez, J., Brun, C., Mirkin, S., & Roux, C. (2015). *XRCE personal language analytics engine for multilingual author profiling* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

*Nowson, Scott, & Oberlander, J. (2007). *Identifying more bloggers: Towards large scale personality classification of personal weblogs* [Conference session]. Weblogs and Social Media, Boulder, CO, United States.

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the Five-Factor Model of personality traits: A meta-analysis. *Journal of Applied Psychology*, *96*(4), 762–773. https://doi.org/10.1037/a0021832

*Ong, V., Rahmanto, A. D., Surhartono, D., Nugroho, A. E., Andangsari, E. W., & Suprayogi, M. N. (2017). *Personality prediction based on Twitter information in Bahasa Indonesia* [Conference session]. 2017 Federated Conference on Computer Science and Information Systems, Prague, Czech Republic. https://doi.org/10.15439/2017F359

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science: Open Science Collobaration. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Orchard, L. J., Fullwood, C., Galbraith, N., & Morris, N. (2014). Individual differences as predictors of social networking. *Journal of Computer-Mediated Communication*, *19*(3), 388–402. https://doi.org/10.1111/jcc4.12068

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*, Article n71. https://doi.org/10.1136/bmj.n71

*Palomino-Garibay, A., Camacho-González, A. T., Fierro-Villaneda, R. A., Hernández-Farias, I., Buscaldi, D., & Meza-Ruiz, I. V. (2015). *A random forest approach for authorship profiling* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

*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), 934–952. https://doi.org/10.1037/pspp0000020

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

*Paudel, A., Bajracharya, B. R., Ghimire, M., Bhattarai, N., & Baral, D. S. (2018). *Using personality traits information from social media for music recommendation* [Conference session]. Computing, Communication and Security (ICCCS), Kathmandu, Nepal.

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology*, *63*(5), 816–824. https://doi.org/10.1037/0022-3514.63.5.816

Pelletier, M. J., Krallman, A., Adams, F. G., & Hancock, T. (2020). One size doesn't fit all: A uses and gratifications analysis of social media platforms. *Journal of Research in Interactive Marketing*, *14*(2), 269–284. https://doi.org/10.1108/JRIM-10-2019-0159

*Peng, K. H., Liou, L. H., Chang, C. S., & Lee, D. S. (2015). *Predicting personality traits of Chinese users based on Facebook wall posts* [Conference session]. 2015 24th Wireless and Optical Communication Conference, Taipei, Taiwan. https://doi.org/10.1109/WOCC.2015.7346106

Peng, Y. X., Zhu, W. W., Zhao, Y., Xu, C. S., Huang, Q. M., Lu, H. Q., Zheng, Q. H., Huang, T. J., & Gao, W. (2017). Cross-media analysis and reasoning: Advances and directions. *Frontiers of Information Technology and Electronic Engineering*, *18*(1), 44–57. https://doi.org/10.1631/FITEE.1601787

*Pervaz, I., Ameer, I., Sittar, A., & Nawab, R. M. A. (2015). *Identification of author personality traits using stylistic features* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

Peter, J., & Valkenburg, P. M. (2006). Research note: Individual differences in perceptions of internet communication. *European Journal of Communication*, *21*(2), 213–226. https://doi.org/10.1177/0267323105064046

*Philip, J., Shah, D., Nayak, S., Patel, S., & Devashrayee, Y. (2019). Machine learning for personality analysis based on big five model. In V. Balas, N. Sharma, & A. Chakrabarti (Eds.), *Data management, analytics and innovation. Advances in intelligent systems and computing* (Vol. 839, pp. 345–355). Springer. https://doi.org/10.1007/978-981-13-1274-8_27

Phua, J., Jin, S. V., & Kim, J. (2017). Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat. *Computers in Human Behavior*, *72*, 115–122. https://doi.org/10.1016/j.chb.2017.02.041

Pineau, J., Vincent-Lamarre, P., Sinha, K., Lariviére, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E., & Larochelle, H. (2021). Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *The Journal of Machine Learning Research*, *22*(1), 7459–7478.

*Ponce-Lopez, V., Chen, B., Oliu, M., Corneanu, C., Clapes, A., Guyon, I., Baro, X., & Escalante, H. J. (2016). *Chalearn lap 2016: First round challenge on first impressions-dataset and results* [Conference session]. 14th European Conference on Computer Vision, Amsterdam, The Netherlands.

*Posadas-Durán, J. P., Markov, I., Gómez-Adorno, H., Sidorov, G., Batyrshin, I., Gelbukh, A., & Pichardo-Lagunas, O. (2015). *Syntactic N-grams as features for the author profiling task* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

*Potash, P., & Rumshisky, A. (2016). *Recommender system incorporating user personality profile through analysis of written reviews* [Conference session]. EMPIRE @RecSys. Boston, MA, United States. https://doi.org/10.1016/j.aqpro.2013.07.003

*Poulston, A., Stevenson, M., & Bontcheva, K. (2015). *Topic Models and n-gram Language Models for author profiling notebook for PAN at CLEF 2015* [Conference session]. CLEF 2015 Labs and Workshops. Notebook Papers, Toulouse, France.

*Pratama, B. Y., & Sarno, R. (2015). *Personality classification based on Twitter text using Naive Bayes, KNN and SVM* [Conference session]. 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, Indonesia. https://doi.org/10.1109/ICODSE.2015.7436992

*Pratama, R. P., & Maharani, W. (2021). *Predicting Big Five personality traits based on Twitter user random forest method* [Conference session]. 2021 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia. https://doi.org/10.1109/ICoDSA53588.2021.9617501

*Principi, R. D. P., Palmero, C., Junior, J. C. S. J., & Escalera, S. (2021). On the effect of observed subject biases in apparent personality analysis from audio-visual signals. *IEEE Transactions on Affective Computing*, *12*(3), 607–621. https://doi.org/10.1109/TAFFC.2019.2956030

*Przybyła, P., & Teisseyre, P. (2015). *What do your look-alikes say about you? Exploiting strong and weak similarities for author profiling* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

*Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, *46*(6), 710–718. https://doi.org/10.1016/j.jrp.2012.08.008

Quan-Haase, A., Martin, K., & McCay-Peet, L. (2015). Networks of digital humanities scholars: The informational and social uses and gratifications of Twitter. *Big Data & Society*, *2*(1), 1–12. https://doi.org/10.1177/2053951715589417

Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of*

*Science, Technology & Society*, *30*(5), 350–361. https://doi.org/10.1177/0270467610380009

*Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft, J. (2012). *The personality of popular facebook users* [Conference session]. ACM 2012 Conference on Computer Supported Cooperative Work, Seattle, WA, United States. https://doi.org/10.1145/2145204.2145346

*Rai, N. (2016). *Bi-modal regression for apparent personality trait recognition* [Conference session]. 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico. https://doi.org/10.1109/ICPR.2016.7899607

Rangel Pardo, F. M., Celli, F., Rosso, P., Potthast, M., Stein, B., & Daelemans, W. (2015). *Overview of the 3rd author profiling task at PAN 2015* [Conference session]. CLEF 2015 Evaluation Labs and Workshop Working Notes Papers, Toulouse, France.

Raudenbush, S. W., Hunter, J. E., & Schmidt, F. L. (2006). Methods of meta-analysis: Correcting error and bias in research findings. *Journal of the American Statistical Association*, *86*(413), 242–244. https://doi.org/10.2307/2289738

*Robles-Granda, P., Lin, S., Wu, X., Martinez, G. J., Mattingly, S. M., Moskal, E., Striegel, A., Chawla, N. V., D'Mello, S., Gregg, J., Nies, K., Mark, G., Grover, T., Campbell, A. T., Mirjafari, S., Saha, K., De Choudhury, M., & Dey, A. K. (2021). Jointly predicting job performance, personality, cognitive ability, affect, and well-being. *IEEE Computational Intelligence Magazine*, *16*(2), 46–61. https://doi.org/10.1109/MCI.2021.3061877

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141–160. https://doi.org/10.1037/met0000300

*Rohit, G. V., Bharadwaj, K. R., Hemanth, R., Pruthvi, B., & Manoj Kumar, M. V. (2020). *Machine intelligence based personality prediction using social profile data* [Conference session]. Smart Systems and Inventive Technology, ICSSIT 2020, Tirunelveli, India. https://doi.org/10.1109/ICSSIT48917.2020.9214175

*Rosen, P. A., McLarty, B. D., Esken, C., Solomon, S., & Taylor, E. (2014). The use of Twitter Profiles to assess personality and hireability. *Academy of Management*, *30*(2), 256–272. https://doi.org/10.5465/ambpp.2014.17496abstract

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. C. Valentine (Eds.), *The handbook of research synthesis* (p. 239). Sage Publications.

Rough, D., de Paoli, S., & Wilson, A. (2021). Commonshare: A new approach to social reputation for online collaborative communities. *Social Science Computer Review*, *41*(1), 4–26. https://doi.org/10.1177/08944393211028191

*Rouse, S. V., & Haas, H. A. (2003). Exploring the accuracies and inaccuracies of personality perception following internet-mediated communication. *Journal of Research in Personality*, *73*(5), 446–467. https://doi.org/10.1016/S0092-6566(03)00025-4

*Rudra, U., Chy, A. N., & Seddiqui, M. H. (2020). *Personality traits detection in Bangla: A benchmark dataset with comparative performance analysis of state-of-the-art methods* [Conference session]. Computer and Information Technology, Proceedings, Dhaka, Bangladesh. https://doi.org/10.1109/ICCIT51783.2020.9392722

*Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How are personality states associated with smartphone data? *European Journal of Personality*, *34*(5), 687–713. https://doi.org/10.1002/per.2309

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*(1), 19–30. https://doi.org/10.1037/1082-989X.13.1.19

Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior*, *27*(5), 1658–1664. https://doi.org/10.1016/j.chb.2011.02.004

*Sarkar, C., Bhatia, S., Agarwal, A., & Li, J. (2014). *Feature analysis for computational personality recognition using Youtube personality dataset* [Conference session]. 2014 ACM multi media on workshop on computational personality recognition, Oriando, FL, United States. https://doi.org/10.1145/2659522.2659528

Scherer, K. R. (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, *40*(2), 191–210. https://doi.org/10.1111/j.1467-6494.1972.tb00998.x

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, *8*(4), 467–487. https://doi.org/10.1002/ejsp.2420080405

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Sage Publications. https://doi.org/10.4135/9781483398105

Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, *62*(1), 97–128. https://doi.org/10.1348/000711007X255327

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. *European Journal of Personality*, *34*(5), 733–752. https://doi.org/10.1002/per.2258

*Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, *8*(9), Article e73791. https://doi.org/10.1371/journal.pone.0073791

*Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., Cristani, M., & Lepri, B. (2017). *What your Facebook profile picture reveals about your personality* [Conference session]. 25th ACM International Conference on Multimedia.

*Segalin, C., Cheng, D. S., & Cristani, M. (2017). Social profiling through image understanding: Personality inference using convolutional neural networks. *Computer Vision and Image Understanding*, *156*, 34–50. https://doi.org/10.1016/j.cviu.2016.10.013

*Segalin, C., Perina, A., Cristani, M., & Vinciarelli, A. (2017). The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits. *IEEE Transactions on Affective Computing*, *8*(2), 268–285. https://doi.org/10.1109/TAFFC.2016.2516994

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. https://doi.org/10.1016/j.paid.2012.10.009

Senn, S. J. (2009). Overstating the evidence—Double counting in meta-analysis and related problems. *BMC Medical Research Methodology*, *9*, Article 10. https://doi.org/10.1186/1471-2288-9-10

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, *21*(4), 217–228. https://doi.org/10.1089/cyber.2017.0384

*Shah, V., & Modi, S. (2021). *Comparative analysis of psychometric prediction system* [Conference session]. Smart Technologies Communication and Robotics, Sathyamangalam, India. https://doi.org/10.1109/STCR51658.2021.9588950

Shaw, H., Brown, O., Hinds, J., Nightingale, S., Towse, J. N., & Ellis, D. (2023). *Describing Ethical Choices In Digital-Data Explorations (DECIDE)–A toolkit*. https://osf.io/preprints/psyarxiv/c8ert/

*Shen, J., Brdiczka, O., Ducheneaut, N., Yee, N., & Begole, B. (2012). *Inferring personality of online gamers by fusing multiple-view predictions* [Conference session]. User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada. https://doi.org/10.1007/978-3-642-31454-4_22

*Shen, J., Brdiczka, O., & Liu, J. (2013). *Understanding email writers: Personality prediction from email messages* [Conference session].

User Modeling, Adaptation, and Personalization: 21th International Conference, UMAP 2013, Rome, Italy. https://doi.org/10.1007/978-3-642-38844-6_29

*Shen, J., Brdiczka, O., & Liu, J. (2015). A study of Facebook behavior: What does it tell about your neuroticism and extraversion? *Computers in Human Behavior*, *45*, 32–38. https://doi.org/10.1016/j.chb.2014.11.067

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

*Sitaraman, G. (2015). *Inferring Big 5 personality from online social networks* [Master's thesis]. University of Washington.

*Skowron, M., Ferwerda, B., Tkalčič, M., & Schedl, M. (2016). *Fusing social media cues: Personality prediction from Twitter and Instagram* [Conference session]. WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web. https://doi.org/10.1007/s11200-011-9001-0

Slawson, N. (2018, April 7). Faceblock campaign urges users to boycott Facebook for a day. *The Guardian*. https://www.theguardian.com/technology/2018/apr/07/faceblock-campaign-urges-users-boycott-facebook-for-one-day-protest-cambridge-analytica-scandal

Smith, S. A., & Watkins, B. (2020). Millennials' uses and gratifications on LinkedIn: Implications for recruitment and retention. *International Journal of Business Communication*, *60*(2), 560–586. https://doi.org/10.1177/2329488420973714

*Solinger, C., Hirshfield, L., Hirshfield, S., Friendman, R., & Leper, C. (2014). *Beyond facebook personality prediction: A multidisciplinary approach to predicting social media users' personality* [Conference session]. Social Computing and Social Media: 6th International Conference, SCSM 2014, Held as Part of HCI International 2014, Heraklion, Crete. https://doi.org/10.1007/978-3-319-07632-4_46

*Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality. *Human-Centric Computing and Information Sciences*, *8*, Article 24. https://doi.org/10.1186/s13673-018-0147-4

*Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(30), 17680–17687. https://doi.org/10.1073/pnas.1920484117

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality*, *34*(5), 613–631. https://doi.org/10.1002/per.2257

*Staiano, J., Pianesi, F., Lepri, B., Sebe, N., Aharony, N., & Pentland, A. (2012). *Friends don't lie—Inferring personality traits from social network structure* [Conference session]. 2012 ACM Conference on Ubiquitous Computing—UbiComp '12, Pittsburg, PA, United States. https://doi.org/10.1145/2370216.2370266

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055. https://doi.org/10.1016/S0895-4356(01)00377-8

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, *115*(11), 2584–2589. https://doi.org/10.1073/pnas.1708290115

*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. https://doi.org/10.1002/per.1935

*Subramaniam, A., Patel, V., Mishra, A., Balasubramanian, P., & Mittal, A. (2016). *Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features* [Conference session]. Computer Vision–ECCV 2016 Workshops, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-319-49409-8_27

*Sulea, O. M., & Dichiu, D. (2015). *Automatic profiling of Twitter users based on their Tweets: Notebook for PAN at CLEF 2015* [Conference session]. CLEF (Working Notes)2, Toulouse, France.

*Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). *Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets* [Conference session]. 2012 11th International Conference on Machine Learning and Applications, Boca Raton, FL, United States. https://doi.org/10.1109/ICMLA.2012.218

*Sun, X., Liu, B., Meng, Q., Cao, J., Luo, J., & Yin, H. (2020). Group-level personality detection based on text generated networks. *World Wide Web*, *23*(3), 1887–1906. https://doi.org/10.1007/s11280-019-00729-2

*Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). Personality predictions based on user behavior on the facebook social media platform. *IEEE Access*, *6*, 61959–61969. https://doi.org/10.1109/ACCESS.2018.2876502

*Tandera, T., Hendro, Suhartono, D., Wongso, R., & Prasetio, Y. L. (2017). Personality prediction system from Facebook users. *Procedia Computer Science*, *116*, 604–611. https://doi.org/10.1016/j.procs.2017.10.016

*Tareaf, R., Alhosseini, S. A., & Meinel, C. (2019). *Facial-based personality prediction models for estimating individuals private traits* [Conference session]. Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 201, Xiamen, China. https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00233

*Tareaf, R., Berger, P., Hennig, P., & Meinel, C. (2020). Cross-platform personality exploration system for online social networks: Facebook vs. Twitter. *Web Intelligence*, *18*(3), 1–17. https://doi.org/10.3233/WEB-20042

*Tareaf, R. B., Alhosseini, S. A., & Meinel, C. (2020). *Does personality evolve? A ten-years longitudinal study from social media platforms* [Conference session]. 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, Exeter, United Kingdom. https://ieeexplore.ieee.org/abstract/document/9443729

*Teijeiro-Mosquera, L., Biel, J. I., Alba-Castro, J. L., & Gatica-Perez, D. (2015). What your face vlogs about: Expressions of emotion and big-five traits impressions in youtube. *IEEE Transactions on Affective Computing*, *6*(2), 193–205. https://doi.org/10.1109/TAFFC.2014.2370044

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*(13), 2113–2126. https://doi.org/10.1002/sim.1461

*Thilakaratne, M., Weerasinghe, R., & Perera, S. (2016). *Knowledge-driven approach to predict personality traits by leveraging social media data* [Conference session]. 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE, USA. https://doi.org/10.1109/WI.2016.0048

Thorndike, R. L., & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin*, *34*(5), 275–285. https://doi.org/10.1037/h0053850

*Tighe, E., & Cheng, C. (2018). *Modeling personality traits of filipino twitter users* [Conference session]. Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, New Orleans, LA, United States.

Timms, H., & Heimans, J. (2018, April 16). Commentary: #DeleteFacebook is just the beginning. Here's the movement we could see next. *Fortune*. https://fortune.com/2018/04/16/delete-facebook-data-privacy-movement/

*Titov, S., Novikov, P., & Mararitsa, L. (2019). *Full-scale personality prediction on VKontakte social network and its applications* [Conference session]. 25th Conference of Open Innovations Associateion (FRUCT), Helsinki, Finland.

*Tomlinson, M., Hinote, D., & Bracewell, D. (2013). *Predicting conscientiousness through semantic analysis of facebook posts* [Conference session]. Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, United States.

*Torfason, R., Agustsson, E., Rothe, R., & Timofte, R. (2016). *From face images and attributes to attributes* [Conference session]. Asian Conference on Computer Vision, Taipei, Taiwan.

Tsikerdekis, M., & Zeadally, S. (2014). Online deception in social media. *Communications of the ACM*, *57*(9), 72–80. https://doi.org/10.1145/2629612

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, *49*, 25–30. https://doi.org/10.1016/j.jrp.2013.12.004

*Tutaysalgir, E., Karagoz, P., & Toroslu, I. H. (2019). *Clustering based personality prediction on Turkish tweets* [Conference session]. 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, BC, Canada. https://doi.org/10.1145/3341161.3343513

Twomey, C., & O'Reilly, G. (2017). Associations of self-presentation on Facebook with mental health and personality variables: A systematic review. *Cyberpsychology, Behavior, and Social Networking*, *20*(10), 587–595. https://doi.org/10.1089/cyber.2017.0247

*Usselmann, H., Ahmad, R., & Siemon, D. (2021). A personality mining system for German Twitter posts with global vectors word embedding. *IEEE Access*, *9*, 165576–165610. https://doi.org/10.1109/ACCESS.2021.3130937

*Vaidhya, M., Shrestha, B., Sainju, B., Khaniya, K., & Shakya, A. (2017). *Personality traits analysis from Facebook data* [Conference session]. 21st International Computer Science and Engineering Conference (ICSEC), Bangkok, Thailand. https://doi.org/10.1109/ICSEC.2017.8443932

Vailshery, L. S. (2022). Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2021, with forecasts from 2022 to 2030. *Statista*. https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

*van de Ven, N., Bogaert, A., Serlie, A., Brandt, M. J., & Denissen, J. J. A. (2017). Personality perception based on LinkedIn profiles. *Journal of Managerial Psychology*, *32*(6), 418–429. https://doi.org/10.1108/JMP-07-2016-0220

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, *47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

van Mil, F. (2020). *Inferring personality from GitHub communication data: Promises & perils* [Master's thesis]. Delft University of Technology.

*van Mil, F. C. J., Rastogi, A., & Zaidman, A. (2021). *Promises and perils of inferring personality on github* [Conference session]. Empirical Software Engineering and Measurement, Bari, Italy. https://doi.org/10.1145/3475716.3475775

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. https://doi.org/10.1037/a0017908

*Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, *87*(1), 123–132. https://doi.org/10.1037/0022-3514.87.1.123

*Verhoeven, B., Daelemans, W., & De Smedt, T. (2013). *Ensemble methods for personality recognition* [Conference session]. Workshop on Computational Personality Recognition: Shared Task, Cambridge, MA, United States. https://doi.org/10.1088/0029-5515/49/6/065035

*Verhoeven, B., Soler Company, J., & Daelemans, W. (2014). *Evaluating content-independent features for personality recognition* [Conference session]. 2014 ACM Multi Media on Workshop on Computational Personality Recognition, Orlando, FL, United States. https://doi.org/10.1145/2659522.2659527

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, *5*(3), 273–291. https://doi.org/10.1109/TAFFC.2014.2330816

*Waggoner, A. S., Smith, E. R., & Collins, E. C. (2009). Person perception by active versus passive perceivers. *Journal of Experimental Social Psychology*, *45*(4), 1028–1031. https://doi.org/10.1016/j.jesp.2009.04.026

*Wald, R., Khoshgoftaar, T., & Sumner, C. (2012). *Machine prediction of personality from Facebook profiles*. Information Reuse and Integration, Las Vegas, NV, United States. https://doi.org/10.1109/IRI.2012.6302998

*Wall, H. J., Kaye, L. K., & Malone, S. A. (2016). An exploration of psychological factors on emoticon usage and implications for judgement accuracy. *Computers in Human Behavior*, *62*, 70–78. https://doi.org/10.1016/j.chb.2016.03.040

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). Sage Publications. https://cir.nii.ac.jp/crid/1370853567712679311

*Wan, D., Zhang, C., Wu, M., & An, Z. (2014). Personality prediction based on all characters of user social media information. In H. Huang, T. Liu, H. P. Zhang, & J. Tang (Eds.), *Social media processing. SMP 2014. Communications in computer and information science* (Vol. 489, pp. 220–230). Springer. https://doi.org/10.1007/978-3-662-45558-6_20

*Wang, H., & Marsella, S. (2017). *Assessing personality through objective behavioral sensing* [Conference session]. Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, United States, San Antonio, TX, United States. https://doi.org/10.1109/ACII.2017.8273590

*Wang, M., Zuo, W., & Wang, Y. (2015). A novel adaptive conditional probability-based predicting model for user's personality traits. *Mathematical Problems in Engineering*, *2015*, Article 472917. https://doi.org/10.1155/2015/472917

*Wang, S., & Chen, X. (2020). Recognizing CEO personality and its impact on business performance: Mining linguistic cues from social media. *Information & Management*, *57*(5), Article 103173. https://doi.org/10.1016/j.im.2019.103173

*Wang, S., Cui, L., Liu, L., Lu, X., & Li, Q. (2020). *Personality traits prediction based on users' digital footprints in social networks via attention RNN* [Conference session]. 2020 IEEE International Conference on Services Computing (SCC), Beijing, China. https://ieeexplore.ieee.org/abstract/document/9284463

*Wang, W., Harari, G. M., Wang, R., Müller, S. R., Mirjafari, S., Masaba, K., & Campbell, A. T. (2018). *Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits* [Conference session]. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. https://doi.org/10.1145/3264951

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, *114*(2), 246–257. https://doi.org/10.1037/pspa0000098

*Wang, Z., Wu, C., Zheng, K., Niu, X., & Wang, X. (2019). SMOTETomek-based resampling for personality recognition. *IEEE Access*, *7*, 129678–129689. https://doi.org/10.1109/ACCESS.2019.2940061

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*(3), 546–558. https://doi.org/10.1037/0022-3514.78.3.546

*Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., Rui, Y., & Ma, W.-Y. (2017). *Beyond the words: Predicting user personality from* *heterogenous information* [Conference session]. Web Search and Data Mining, Cambridge, United Kingdom. https://doi.org/10.1145/3018661.3018717

*Wei, X. S., Zhang, C. L., Zhang, H., & Wu, J. (2018). Deep bimodal regression of apparent personality traits from short video sequences. *IEEE Transactions on Affective Computing*, *9*(3), 303–315. https://doi.org/10.1109/TAFFC.2017.2762299

Wells, G., Shea, B., O'Connell, D., & Peterson, J. (2000). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. Ottawa Hospital Research Institute.

*Weren, E. R. D. (2015). *Information retrieval features for personality traits* [Conference session]. CEUR Workshop Proceedings, Toulouse, France.

*Werlen, L. (2015). *Statistical Learning Methods for Profiling Analysis Notebook for PAN at CLEF 2015* [Conference session]. CLEF 2015 Labs and Workshops. Notebook Papers, Toulouse, France.

Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research*, *16*(4), 362–369. https://doi.org/10.1108/QMR-06-2013-0041

Whittle, H., Hamilton-Giachritsis, C., Beech, A., & Collings, G. (2013). A review of online grooming: Characteristics and concerns, *Aggression and Violent Behavior*, *18*(1), 62–70. https://doi.org/10.1016/j.avb.2012.09.003

Whitty, M. T. (2018). Do you love me? Psychological characteristics of romance scam victims. *Cyberpsychology, Behavior, and Social Networking*, *21*(2), 105–109. https://doi.org/10.1089/cyber.2016.0729

Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, *3*(1), 94–123. https://doi.org/10.1177/2515245919885611

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), 395–398. https://doi.org/10.1089/cyber.2017.0652

Wright, K. (2016). Communication in health-related online social support groups/communities: A review of research on predictors of participation, applications of social support theory, and health outcomes. *Review of Communication Research*, *4*, 65–87. https://doi.org/10.12840/issn.2255-4165.2016.04.01.010

*Wu, T., & Zheng, Y. (2019). Is impression management through status updates successful? Meta-accuracy and judgment accuracy of big five personality traits based on status updates from social network sites in china. *Frontiers in Psychology*, *10*, Article 1192. https://doi.org/10.3389/fpsyg.2019.01192

*Xianyu, H., Xu, M., Wu, Z., & Cai, L. (2016). *Heterogeneity-entropy based unsupervised feature learning for personality prediction with cross-media data* [Conference session]. IEEE International Conference on Multimedia and Expo, Seattle, WA, United States. https://doi.org/10.1109/ICME.2016.7552980

*Xiong, X., Filippone, M., & Vinciarelli, A. (2016). *Looking good with Flickr faves: Gaussian processes for finding difference makers in personality impressions* [Conference session]. MM'16: Proceedings of the 2016 ACM Multimedia Conference. https://doi.org/10.1145/2964284.2967253

*Xu, R., Frey, R. M., & Ilic, A. (2016). *Individual differences and mobile service adoption: An empirical analysis* [Conference session]. 2016 IEEE 2nd International Conference on Big Data Computing Service and Applications, BigDataService 2016, Oxford, United Kingdom. https://doi.org/10.1109/BigDataService.2016.15

*Xue, D., Hong, Z., Guo, S., Gao, L., Wu, L., Zheng, J., & Zhao, N. (2017). Personality recognition on social media with label distribution learning. *IEEE access*, *5*, 13478–13488. https://doi.org/10.1109/ACCESS.2017.2719018

*Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., Zhong, X., & Sun, J. (2018). Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence*, *48*(11), 4232–4246. https://doi.org/10.1007/s10489-018-1212-4

Yakoub, F., Zein, M., Yasser, K., Adl, A., & Hassanien, A. E. (2015). Predicting personality traits and social context based on mining the smartphones SMS data. In A. Abraham, X. Jiang, V. Snášel, & J. S. Pan (Eds.), *Intelligent data analysis and applications. Advances in intelligent systems and computing* (Vol. 370, pp. 511–521). Springer, Cham. https://doi.org/10.1007/978-3-319-21206-7_44

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

*Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *National Academy of Sciences*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

*Yu, J., & Markov, K. (2017). *Deep learning based personality recognition from Facebook status updates* [Conference session]. Awareness Science and Technology, Taichung, Taiwan. https://doi.org/10.1109/ICAwST.2017.8256484

*Yuan, C., Wu, J., Li, H., & Wang, L. (2018). *Personality recognition based on user generated content* [Conference session]. Service Systems and Service Management, Hangzhou, China. https://doi.org/10.1109/ICSSSM.2018.8465006

Zarouali, B., Dobber, T., de Pauw, G., & de Vreese, C. (2020). Using a personality-profiling algorithm to investigate political microtargeting: Assessing the Persuasion Effects of personality-tailored ads on social media. *Communication Research*, *49*(8), 23–31. https://doi.org/10.1177/0093650220961965

*Zhang, L., Zhao, L., Zhang, X., Kong, W., Sheng, Z., & Lu, C. T. (2018). *Situation-based interpretable learning for personality prediction in social media* [Conference session]. Big Data (Big Data), Seattle, WA, United States.

Zhang, Q. (2022). Meta-analysis of correlation coefficients: A cautionary tale on treating measurement error. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000498

*Zheng, H., & Chunhua, W. (2019). *Predicting personality using Facebook status based on semi-supervised learning* [Conference session]. Machine Learning and Computing, Zhuhai China. https://dl.acm.org/doi/abs/10.1145/3195106.3195124?casa_token=d7_2QAkGDrIAAAAA:LQzcyw9-lJZrGu0GP0X5L3-C-2IZ3mZdutjl6tWTCiTg5TDBjGZJJUftALzaFa0CYyiVtCiKYZseZw

*Zhong, X. F., Guo, S. Z., Gao, L., Shan, H., & Xue, D. (2018). *A general personality prediction framework based on Facebook profiles* [Conference session]. Proceedings of the 2018 10th International Conference on Machine Learning and Computing. Macau, China. https://dl.acm.org/doi/abs/10.1145/3195106.3195124?casa_token=d7_2QAkGDrIAAAAA:LQzcyw9-lJZrGu0GP0X5L3-C-2IZ3mZdutjl6tWTCiTg5TDBjGZJJUftALzaFa0CYyiVtCiKYZseZw

Zhuang, F., Luo, P., Xiong, H., Xiong, Y., He, Q., & Shi, Z. (2010). Cross-domain learning from multiple sources: A consensus regularization perspective. *IEEE Transactions on Knowledge and Data Engineering*, *22*(12), 1664–1678. https://doi.org/10.1109/TKDE.2009.205

*Ziyi, Y., Jun, S., Daxiang, H., Tong, W., & Zhili, X. (2016). Personality as a metric for topic models on social networks. *Journal of High Speed Networks*, *22*(2), 169–176. https://doi.org/10.3233/JHS-160540

Received January 15, 2020
Revision received January 4, 2024
Accepted January 23, 2024 ■