European Journal of Personality, *Eur. J. Pers.* **33**: 197–213 (2019)
Published online 7 March 2019 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.2196

The Effect of Information Quantity on Distinctive Accuracy and Normativity of Personality Trait Judgments

SHEHEREZADE L. KRZYZANIAK[1]*, DOUGLAS E. COLMAN[1†], TERA D. LETZRING[1], JENNIFER S. MCDONALD[1] and JEREMY C. BIESANZ[2]

[1]*Department of Psychology, Idaho State University, Pocatello, ID USA*
[2]*Department of Psychology, University of British Columbia, Vancouver, British Columbia Canada*

Abstract: Information quantity is an important moderator of personality judgment accuracy. Some evidence suggests that the amount of available information is positively related to accuracy. The current study utilized the social accuracy model to investigate the effects of differences in thin slices of information quantity on the distinctive accuracy and normativity of personality trait judgments. It was hypothesized that distinctive accuracy and normativity would increase as information quantity increased. Participants were 431 individuals who participated in an online study that varied the length of stimulus target observations (30 seconds, 1 minute, 3 minutes, and 5 minutes), after which judges rated targets using other-report measures of the Big Five personality traits. For all traits combined, significant levels of accuracy were found for all observation lengths, but distinctive accuracy and normativity did not increase as video length increased. Findings varied for individual traits. For distinctive accuracy, there was a linear increase with information quantity for Extraversion and a non-linear relationship for Conscientiousness, while there was a linear decrease for Openness. For normativity, there was a linear increase with information quantity for Agreeableness and a non-linear relationship for Conscientiousness. There are important differences in how observation length affects distinctive accuracy and normativity for different personality traits. © 2019 European Association of Personality Psychology

Key words: personality judgment accuracy; information quantity; social accuracy model; distinctive accuracy; normativity

Making judgments of other people is a common task in everyday interactions. When meeting someone new, it is helpful to be able to accurately figure out what that person is like, as this information can influence our perceptions and interactions. Many aspects play a role in the level of accuracy with which a judge (the person making judgments) evaluates a target (the person being judged), such as characteristics and behaviour of judges, expressiveness of targets, and the type and amount of information provided by the target (Funder, 1995; Letzring & Funder, 2018). The amount of information is referred to as *information quantity*, and this moderator is the focus of the current investigation. When meeting an individual for the first time, one does not always have the luxury of hours or days to get to know that person, and in situations such as a job interview or speed dating, there is often a very small window of time before an important initial decision about a person must be made. The positive relationship between information quantity and accuracy when there are large differences in the amount of information available to judges (i.e. strangers vs years of acquaintance) is supported by previous research (Beer, in press; Biesanz, West, & Millevoi, 2007; Colvin & Funder, 1991; Funder & Colvin, 1988). There is also a research supporting the idea that accuracy is possible for certain traits with just seconds or minutes of information (Carney, Colvin, & Hall, 2007) and that accuracy improves as the length of exposure increases from 5 to 30 minutes (Blackman & Funder, 1998; Letzring, Wells, & Funder, 2006). However, it is less clear whether accuracy increases over smaller differences in information quantity (such as between 30 seconds and 1 minute), and how accuracy changes over time when the components of distinctive accuracy and normativity is examined. Distinctive accuracy reflects the degree of match between the judges' ratings of targets and what those targets are actually like, whereas normativity (also termed stereotype accuracy or normative accuracy) reflects the degree of match between the judges' ratings of targets and what the average person is like. The current study examined how these components of accuracy were affected by differences in exposures to video recording of targets of durations between 30 seconds and 5 minutes.

MODELS OF ACCURACY

Realistic accuracy model

The realistic accuracy model (RAM; Funder, 1995, 1999; Letzring & Funder, in press) is often used to describe the

*Correspondence to: Sheherezade L. Krzyzaniak, Department of Psychology, Idaho State University, 921 S. 8th Ave. Stop 8112, Pocatello, ID 83209, USA. E-mail: krzysheh@isu.edu
†Current Address: University of Wisconsin - La Crosse, La Crosse, WI USA.

This article earned Open Data and Open Materials badges through Open Practices Disclosure from the Center for Open Science: https://osf.io/tvyxz/wiki. The data and materials are permanently and openly accessible at https://osf.io/mrt8f/. Author's disclosure form may also be found at the Supporting Information in the online version.

Handling editor: Mitja Back
Received 7 October 2017
Revised 1 February 2019, Accepted 2 February 2019

process that takes place when an accurate judgment is made. This model describes a four-stage process in which each stage must be successfully completed for an accurate judgment to be possible. The process includes the stages of relevance, availability, detection, and utilization. First, cues must exist that are *relevant* to the trait being judged. For example, if Extraversion was the trait of interest, there must be thoughts or behaviours present that are related to a target's level of Extraversion. Next, the target must make those cues *available* or external within the environment. In the example of Extraversion, the target must not only have thoughts related to his or her level of Extraversion but must also externalize those cues in a way that is noticeable by others. Then, the judge must *detect* the relevant and available cues displayed by the target and correctly *utilize* those cues to make a judgment about the trait of interest. For example, the judge must conclude that a behaviour is related to the trait of Extraversion, as well as how extraverted a target is based on that cue.

In addition to describing the process by which accurate judgments are possible, RAM also describes four moderators of accuracy: the good target, good judge, good trait, and good information. The *good target* reflects that some people are more easily and accurately judged than others (Colvin, 1993; Human & Biesanz, 2013; Mignault & Human, in press). The *good judge* reflects that some people tend to make more accurate judgments than others (Colman, in press; Colman, Letzring, & Biesanz, 2017; Kolar, 1996; Letzring, 2008; Vogt & Colvin, 2003). The *good trait* reflects that some personality traits are more easily judged than others (Beer & Watson, 2010; Borkenau & Liebler, 1992; Funder & Dobroth, 1987; John & Robins, 1993; Krzyzaniak & Letzring, in press; Naumann, Vazire, Rentfrow, & Gosling, 2009). *Good information* has two aspects: *quantity* or how much information is available to the judge and *quality* or how relevant the information is to the trait being judged (Beer, in press; Blackman & Funder, 1998; Letzring et al., 2006; Letzring & Human, 2014). The current paper focuses on the quantity of information and utilizes the framework of RAM when conceptualizing how information quantity is expected to affect accuracy. The prediction derived from RAM is that accuracy will increase as information quantity, in the form of the length of recorded interactions that are observed by judges, increases. In the current study, video lengths were 30 seconds, 1 minute, 3 minutes, and 5 minutes, and it was predicted that accuracy would increase across these lengths.

Person model

Another model of personality trait judgment is PERSON (which stands for personality, error, residual, stereotype, opinion, and norm; Kenny, 2004). The components of the PERSON model represent sources of information that people can use when making judgments of others. PERSON is primarily a model of consensus or how closely two or more judges agree in their ratings of a target (Kenny, Albright, Malloy, & Kashy, 1994). Ratings with high consensus may also be highly accurate, but it is possible to have high consensus without high accuracy in the case that people agree with each other but are all inaccurate. Kenny (2004) predicted that when personality is used as the accuracy criterion and the parameters are set based on previous research results, accuracy will increase with information quantity, and this increase will be especially apparent at low levels of acquaintance or within the first 10 or so behaviours of the target (p. 272).[1] As the current project uses personality as the accuracy criterion and examines differences in information quantity based on a range of 30 seconds to 5 minutes of observation, the prediction of the current project that there will be a positive relationship between accuracy and observation length is supported by PERSON.

COMPUTATIONS OF DIFFERENT TYPES OF ACCURACY

Accuracy has been calculated in many ways over the years, which has sometimes led to some confusion about what accuracy scores mean and whether different kinds of accuracy scores should be compared with each other. The information that follows describes the methods that are currently used to compute accuracy and what the accuracy scores derived from each method mean. These types of accuracy will be referred to in the subsequent discussion of the existing research that has examined the relationship between information quantity and accuracy.

Item-level correlations

Accuracy has been traditionally calculated in two main ways within most of the existing literature (Allik, Borkenau, Hrebícková, Kuppens, & Realo, 2015; Hall et al., 2018). The first type of analysis uses *item-level correlations*, in which ratings on a single item across many judge–target pairs are correlated with the accuracy criteria for that specific item. Item-level correlations are sometimes referred to as *trait accuracy* when scores for a single trait are correlated. Trait accuracy reflects how well a set of judges is able to order a set of targets on a given trait or how well individual judges are able to order a set of targets on a single trait (Hall et al., 2018).

Profile correlations

The second traditional approach of calculating accuracy is with *profile correlations*, in which ratings from judges on several items for a single target are correlated with the accuracy criteria for that same set of items. Profile correlations reflect the similarity in the ordering of the judges' ratings and the accuracy criteria for several items or traits, for each judge–target pair. When judges rate multiple targets, accuracy scores can be averaged across the targets to derive a single accuracy score for each judge, and likewise when targets

[1]Changes in the parameters of the model can result in predictions of no relationship, and even a negative relationship, between accuracy and information quantity. However, the prediction of a positive relationship is based on existing data, and this is most likely to apply to the current study.

DOI: 10.1002/per

are rated by multiple judges, accuracy scores can be averaged across the judges to derive a single accuracy score for each target. It is possible to use the profile correlation approach to look at accuracy for specific traits by using only items that assess a given trait in each correlation.

Components of accuracy

Early accuracy research primarily looked at accuracy as a whole (e.g. Letzring et al., 2006), but when using profile correlations, it is possible to separate accuracy into components, with the most common components being distinctive accuracy and normativity (Furr, 2008). Distinctive (or differential) accuracy is consistent with what lay people typically think accuracy means, in that it reflects the degree of match between the judges' ratings of targets and what those targets are actually like. In other words, judges score higher in distinctive accuracy when they accurately identify how targets differ from the average person and from each other. Normativity (or normative accuracy or stereotype accuracy) reflects the degree of match between the judges' ratings of targets and what the average person is like. In other words, judges score higher on normativity when they rate targets in line with what the average person is like. The normative profile is highly favourable, and for this reason, high normativity also reflects highly favourable ratings (Rogers & Biesanz, 2015). Some recent research has focused on a componential approach when looking at accuracy and has demonstrated the importance and benefit of doing so, in that results often differ based on whether distinctive accuracy or normativity is considered (Biesanz, 2010; Biesanz & Human, 2010; Colman et al., 2017; Furr, 2008; Letzring, 2015; Letzring & Human, 2014).

Interpretations of accuracy scores

It has been suggested that profile and item-level correlations address different questions about how judges are accurate (Hall et al., 2018). Recall that profile correlations reflect how accurately a judge is able to order a set of items for a given target, whereas item-level correlations reflect how accurately a set of judges is able to order a set of targets on a given item. Although the processes of comparing several traits within a single person and comparing many people on a single trait could conceptually be quite different, data are typically collected in the same way regardless of the analysis that is used, with judges being asked to consider and rate one target at a time without explicitly comparing targets. When these two types of analyses are used to examine the same data set, the results for distinctive profile accuracy and item-level accuracy are the same if the data (which must be complete and balanced in that the variance of the predictor is the same for all units in the analysis) are standardized in the same way for both analyses, because these analyses simply compute average correlations in different ways (Allik et al., 2015). The average distinctive profile accuracy is the same as the average item-level accuracy (across items) when calculated as unstandardized relationships and weighting each estimate by its precision (Biesanz, 2018) when the data are centred both within items and within targets. This relationship does not hold for normative accuracy or overall profile accuracy, which incorporate mean-level information that is removed when calculating accuracy at the item level.

Profile distinctive accuracy is the weighted average of the item-level accuracies for the items that comprise the profile. For unstandardized analyses, the specific weights are a function of the variance across targets on the validity measure (see Biesanz, 2018, for more discussion, a detailed example, and open-access data). For correlational analyses, a more complex double standardization is needed to show the equivalence of profile correlation and item-level correlation approaches (Allik et al., 2015). Thus, this average accuracy score represents both the average level of accuracy in discerning traits within individuals (e.g. Is Jane more reliable than talkative?), as well as the average level of accuracy in being able to distinguish individual differences (e.g. Is Jane more reliable than Jake?) across the different items/traits examined in the profile.

Social accuracy model

Integrating these various accuracy metrics, more recent research has utilized multivariate approaches to the conceptualization and estimation of accuracy. In particular, a third way of conceptualizing and estimating accuracy that more recent research has utilized is the social accuracy model (SAM; Biesanz, 2010, 2019). This multilevel model can be used to separate accuracy into the components of distinctive accuracy and normativity and is a more powerful analysis than a correlational approach (Biesanz, 2010; Cronbach, 1955; Furr, 2008). If a multilevel model approach is used to simultaneously examine both distinctive accuracy and normativity, it can provide a more in-depth and powerful explanation of how accuracy is influenced by exposure to different amounts of information (operationalized in the current study in terms of length of video observation) about the targets. The present manuscript examines distinctive accuracy (average accuracy across all traits/items) as a function of information, as well as distinctive accuracy within each of the Big Five traits (average within-target accuracy for items on a specific trait). This latter measure assesses the ability of perceivers to accurately judge deviations around a targets' mean trait level.

INFORMATION QUANTITY AND ACCURACY

Research investigating the role of information quantity and accuracy has been diverse in the types of questions posed, the methodology used to answer those questions, and—as already described—the computation of accuracy itself. Research has investigated information quantity in terms of seconds, minutes, and even years of information (Biesanz et al., 2007; Carney et al., 2007; Letzring et al., 2006). Some studies utilize a third-party-observation approach in which participants watch videos of individuals that span a few seconds to several minutes to a few hours, while other studies have participants engage in interactions with strangers that range from several minutes to a few hours. Other studies

DOI: 10.1002/per

have investigated real-life acquaintanceships to study longer term relationships that go beyond just a few hours of interaction to span a semester or many years (Biesanz et al., 2007; Brown & Bernieri, 2017; Kenny et al., 1994).

Large differences in information quantity

It is generally accepted that accuracy is higher when more information about a target is available to the judge, which is referred to as the *acquaintanceship effect*. This reflects the finding that judgments of personality traits tend to be more accurate when people are judging targets with whom they have been acquainted for an extended period of time in comparison with strangers or people with whom they have only had brief interaction with or exposure to (Biesanz et al., 2007; Blackman & Funder, 1998; Colvin & Funder, 1991; Funder & Colvin, 1988). According to RAM, information quantity moderates accuracy, and research consistently supports this prediction when there are large differences in levels of acquaintanceship between the more and less acquainted groups. Higher levels of self-other agreement, or how much the ratings of a judge agree with the self-ratings of the target, have been consistently found among natural acquaintances compared with strangers who have only limited interaction with or exposure to a target (Blackman & Funder, 1998; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Colvin & Funder, 1991; Funder, 1995; Funder & Colvin, 1988). A meta-analysis that included accuracy as indexed with item-level correlations indicated that judges who have higher levels of physical intimacy with targets (with family members being the highest and strangers the lowest) achieved higher levels of self-other agreement and that this was strongest for the traits of Neuroticism and Openness and weakest for Extraversion (Connelly & Ones, 2010).[2] Among natural acquaintances who had known each other for a few months to many years, length of acquaintance was positively related to self-other agreement calculated as distinctive profile correlations but not to raw profile accuracy or normativity (Biesanz et al., 2007, Study 1). In a replication, similar findings emerged, but length of acquaintanceship was negatively related to normativity (Biesanz et al., 2007, Study 2). Overall, there is strong support for increases in accuracy when there are large differences in the amounts of acquaintanceship.

Smaller differences in information quantity

When differences in the amounts of information quantity are across shorter timespans, findings are mixed. A longitudinal study, in which accuracy was assessed with Kenny's SOREMO program that calculates correlations between self-ratings and the average of judge ratings, did not find consistent increases in self-other agreement from 2 weeks to 8 months for new dormmates (Park, Kraus, & Ryan, 1997), but a longitudinal study of previously unacquainted female college roommates found that self-other agreement increased across acquaintanceships of 2 to 15 weeks for all of the Big Five traits, with significant changes for the traits of Openness and Agreeableness (Kurtz & Sherker, 2003).[3] Another longitudinal study assessed accuracy after one versus seven 20-minute group interactions and found an increase in accuracy for all traits except Extraversion (Paulhus & Reynolds, 1995). Yet another longitudinal study compared self-other agreement of judgments based on no interaction, one 5-minute conversation, and weekly conversations across 10 weeks (totaling about 30 hours of interaction) and found that self-other agreement increased for all traits when comparing ratings before any interaction to ratings after 10 weeks of interaction (Brown & Bernieri, 2017). The trait with the largest increase in self-other agreement over time was Extraversion, although this was the only trait that did not have a greater level of change between 5 minutes and 10 weeks than between no interaction and 5 minutes.

Research has also experimentally manipulated information quantity to examine the effect on accuracy. This has been carried out in a couple of ways, one of which is varying the lengths of interactions or video observations. This methodology was used in the current study, and therefore, results from these studies are the main basis of the hypotheses for the current study. One study found that judgments following 25- to 30-minute observations of targets reached higher levels of self-other agreement, based on raw profile correlations, than judgments following only 5- to 10-minute observations, and this effect was driven by the most visible items and items related to Extraversion and Neuroticism (Blackman & Funder, 1998). Similarly, judgments using the 100 items of the California Adult Q-sort following 50-minute or 3-hour interactions resulted in higher accuracy based on profile correlations than judgments made without prior interaction, although accuracy did not differ for the 50-minute and 3-hour interactions (Letzring et al., 2006).

When manipulating even shorter quantities of exposure to targets, support for the positive relationship between information quantity and accuracy is somewhat mixed. Judgments based on observing videos that ranged from thin slices of 15 seconds to 5 minutes found a positive relationship between video length and item-level correlations based on trait scores, and this effect was driven by Extraversion and Agreeableness (Carney et al., 2007). A different study found that judgments of 5-minute video observations obtained higher levels of distinctive accuracy and normativity (in terms of self-other agreement based on profile correlations) compared with judgments of 30-second video observations (Human, Jackson, & Biesanz, 2008). On the other hand, self-other agreement was not significantly related to increases in information quantity for judgments of 45-second observations compared with judgments of still photographs (Beer & Watson, 2010). Finally, a meta-analysis of 44 studies with more diverse outcomes than just self-other agreement or accuracy (e.g. outcomes related to clinical psychology, social

[2]Please note that this study represents a different way of examining the good information moderator that is less dependent on the length of acquaintance and instead combines information quantity and quality. Duration of acquaintance was coded, which is more similar to the current project, but its relation with self-other agreement was not examined in the meta-analysis.

[3]It is not clear from the research report how accuracy was computed, but it is likely to have been with item-level correlations.

DOI: 10.1002/per

psychology, or deception detection) did not find differences in accuracy between thin-slice observations of 30 seconds and 5 minutes (Ambady & Rosenthal, 1992). In summary, it is unclear whether information quantity moderates accuracy when observation times are relatively short, although previous research indicates that thin slices of information quantity may have more impact on certain traits than others.

CURRENT STUDY AND HYPOTHESES

It was hypothesized that observation length would be positively related to both distinctive accuracy and normativity of judgments of all Big Five traits combined. These hypotheses were identified prior to data analysis but were not preregistered in a formal way. Based on the findings of studies with methodologies similar to the current study,[4] we predicted that distinctive accuracy would be positively related to video length. This prediction is consistent with what would be derived from RAM and PERSON, as noted in the Models of Accuracy section. The prediction from RAM is straightforward, in that judgments based on more information are expected to have higher levels of accuracy. The prediction based on PERSON is more complex and depends on how the parameters of the model are set. When the parameters were based on previous research findings, information quantity was expected to be related to accuracy at low levels of acquaintance, which is consistent with the methodology in the current study.

We also predicted that normativity would be positively related to video length. Only one study has examined the relationship between normativity and information quantity, and either no change or a decrease in normativity was found over several years of acquaintance (Biesanz et al., 2007). Although normativity should theoretically decrease over time, this theory was based on much longer lengths of acquaintance than were used in the current study (months and years vs minutes), and at a very short acquaintance, an increase in stereotype accuracy was predicted (Biesanz et al., 2007, p. 124; Kenny, 1991). The current study looks at only very early acquaintance, so the results of the previous study are unlikely to generalize to the current study. Therefore, it is plausible that normativity does not necessarily decrease until more acquaintanceship has been established (Biesanz et al., 2007; Funder, 1999). Also, normativeness contributes largely to overall accuracy, which has been examined in several studies, and most of the studies that use methodology similar to the current study support the positive relationship between accuracy and information quantity. An alternative hypothesis is that normativity will decrease over time from 30 seconds to 5 minutes, as judges should have to rely less on stereotypes when they are exposed to more individuating cues and are able to more accurately judge how targets are unique. This alternative hypothesis was not formulated prior to data analysis and therefore is not the focus of the current project.

The current study used the SAM to estimate levels of distinctive accuracy and normativity. This research addresses a gap in the literature by examining how two components of accuracy—distinctive accuracy and normativity—are affected by variation in information quantity across thin-slice exposures. The purpose of the current study was to examine the effects of varying information quantity for target stimulus exposures between 30 seconds and 5 minutes on distinctive accuracy and normativity.

METHOD

Adhering to the 21 word solution outlined by Simmons, Nelson, and Simonsohn (2012), 'we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study'. The materials, data, and script for this study can be found on the Open Science Framework (https://osf.io/mrt8f/).

Participant–judges

Participants included 431 individuals (66.36% female) from 46 states in the USA between the ages of 18 and 78 (M_{age} = 36.27; SD_{age} = 12.10). The states with the largest representation included California, Florida, Texas, and New York. Ethnicity was 74.7% Caucasian, 8.35% African American, 6.03% Asian American, 4.64% biracial, and 6.28% other. Participants were recruited for an online study through Amazon's Mechanical Turk (MTurk) and were randomly assigned to one of four conditions, in which the length of the video observations differed (30 seconds, 1 minute, 3 minutes, or 5 minutes). Sample size was based on the goal of having at least 100 participants in each condition. This allowed for an examination of variability of accuracy scores across judges in each condition, which was the original purpose of the project for which these data were collected.[5] This sample size was also sufficient to detect a relatively small effect size (f = .16) with power of .80 and a Type I error rate of α = .05 using a one-way analysis of variance (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007). Having power to detect a small effect was an important feature of this study, as the differences in observation lengths were small, and therefore, any differences in accuracy were also likely to be small.

Accuracy criterion of the targets

Self-reports of personality usually include some bias due to the evaluativeness of some traits and the typical desire to present one's self in a positive manner (Funder, 1995; John & Robins, 1993; Klonsky, Oltmanns, & Turkheimer, 2002; Letzring, 2008; Vazire, 2010). As a result of this bias, it

[4]By this, we mean studies that experimentally manipulated information quantity across short amounts of time by having judges observe videos of different lengths and studies that used overall and distinctive profile correlations and item-level correlations to assess accuracy.

[5]See Letzring and Colman (2018).

DOI: 10.1002/per

has been argued that self-reports are not the best estimate of a person's 'true' personality. Similarly, informant reports, typically solicited from close-known and well-known acquaintances, are also prone to some positivity bias (e.g. Hollander, 1956; Klonsky et al., 2002; Leising, Erbs, & Fritz, 2010). Specifically, acquaintances nominated by targets tend to see the targets positively and with redundancy in regard to the self-reports. While two potentially biased reflections of a target's 'true' personality do not aggregate to be an unbiased representation, combining multiple ratings is almost certainly better than using only one or the other.

The accuracy criterion for each target in this study was composed of self-report ratings of personality as well as ratings from one or two acquaintances who had known the target for at least 6 months (ratings from two acquaintances were available for six targets, and ratings from one acquaintance were available for two targets). Both the target and their acquaintances filled out the same measure of personality, although two different measures of personality were used because videos from four previous studies were used as the stimuli in the current study. These measures were the Big Five Inventory (BFI; John, Naumann, & Soto, 2008) and the International Personality Item Pool 300-item version of the NEO-PI-R (IPIP-NEO-PI-R; Goldberg, 1999; IPIP-NEO-PI domains, n.d.; John, Hampson, & Goldberg, 1991). The BFI is a 44-item measure and has adequate internal reliability (αs = .75–.80) and test–retest reliability over a period of 3 months (*r*s = .80–.90). The original 300-item version of the IPIP-NEO-PI-R has an adequate average reliability across the five dimensions (α = .90; Goldberg, 1999). The two acquaintance ratings were averaged before being averaged with the self-report, so that both types of ratings have equal weight in the accuracy criterion.

Video selection

Video clips all began at the same time-point, regardless of condition, with the only difference being the total length of the video. This means that all participants were exposed to the same targets and a minimum of the same opening 30 seconds of each video. Videos of one female and one male target from each of four existing video libraries were used as stimulus materials in the current study, for a total of eight targets. The targets ranged in age from 19 to 30 years (M = 23.25, SD = 4.29). Two videos contained footage of three unacquainted individuals engaged in an unstructured interaction; one video contained footage of two unacquainted individuals who were instructed to discuss their behaviours across various situations (e.g. when with family, at social events); one video contained footage of two unacquainted individuals who were instructed to engage in a set of behaviours (e.g. reading a poem out loud, playing Jenga); two videos were of an individual engaging in a mock job interview; and two videos contained footage of two unacquainted individuals discussing their hobbies. When more than one person was visible in the video, judges were instructed to focus on a specified person. Targets were also chosen so that there was variability in the levels of the Big Five personality traits, ego-control and ego-resiliency, happiness or satisfaction with life, dominance, and self-reported state affect following two of the interactions. See Table 1 for descriptive statistics for the Big Five traits based on the accuracy criteria. These videos were selected to create highly generalizable findings due to the variation in the targets and situations. Descriptive statistics for the judgments of each trait can also be found in Table 1.

Personality assessment

After viewing each video, participant–judges completed either the BFI or the IPIP-NEO-PI-R, depending on which measure was used to create the accuracy criterion for the specific target. The judges used a condensed version of the IPIP-NEO-PI-R that consisted of 60 items (two per facet) to reduce the length of time necessary to make judgments and to keep the number of items between the two measures of personality comparable.

Procedures

Participation was voluntary through the MTurk website, and a 50-cent compensation was offered as an incentive. Only those who correctly answered at least 80% of attention checks (an example prompt is 'make sure to select *agree a little* to this item' embedded in a matrix of items) and completed at least 80% of the procedure were remunerated and subsequently included in the data analyses. This stipulation was clearly stated in the MTurk posting and the informed consent document. Consenting participants were randomly assigned through an online data collection website to one of the four video length conditions (30 seconds, 1 minute, 3 minutes, or 5 minutes). All participants were informed that they would watch eight videos no more than 5 minutes in length each, but the specific lengths for each condition were not disclosed in an

Table 1. Descriptive statistics of targets and judgments

	Mean	Population *SD*	Range
Accuracy criteria of targets			
Extraversion	3.59	0.78	2.25–4.47
Agreeableness	4.01	0.47	3.22–4.56
Conscientiousness	3.71	0.42	3.22–4.42
Neuroticism	2.62	0.67	1.75–3.81
Openness to Experience	3.55	0.32	3.02–3.93
Ratings from judges			
Extraversion	2.40–4.00	0.57–0.91	1.00–5.00
Agreeableness	3.38–3.93	0.47–0.86	1.08–5.00
Conscientiousness	3.21–3.97	0.60–0.75	1.00–5.00
Neuroticism	2.16–2.98	0.50–0.72	1.00–5.00
Openness to Experience	2.71–3.35	0.38–0.69	1.00–5.00

Note: The population *SD* for the targets was used to describe the eight targets used in the stimulus materials. Descriptives for ratings from judges were computed for each target, so ranges of the values are reported. *SD*, standard deviation.

DOI: 10.1002/per

effort to discourage selection effects.[6] When videos contained more than one individual, participants received instructions regarding which individual to focus on. After viewing each video, participants filled out an other-report assessment of personality for the target in the video using either the BFI or the IPIP-NEO-PI-R, depending on the accuracy criterion available for that target.[7] Finally, participants completed a basic demographic measure.

Data analyses

The SAM (Biesanz, 2010), a cross-classified multilevel model, has been increasingly used within personality judgment research to simultaneously examine both distinctive accuracy and normativity (e.g. Human & Biesanz, 2011; Letzring & Human, 2014; Rogers & Biesanz, 2014). The base model is expressed as follows [Equations (1.1) and (1.2)]:

$$Y_{ijk} = \beta_{0ij} + \beta_{1ij}\, TCrit_{jk} + \beta_{2ij}\, Norm_k + \varepsilon_{ijk} \tag{1.1}$$

$$\begin{aligned} \beta_{0ij} &= \gamma_{00} + u_{0i} + u_{0j} \\ \beta_{1ij} &= \gamma_{10} + u_{1i} + u_{1j} \\ \beta_{2ij} &= \gamma_{20} + u_{2i} + u_{2j} \end{aligned} \tag{1.2}$$

Under this model, Y_{ijk} is judge i's rating of target j (the judge–target pair) on item k of the judgment measure. $TCrit_{jk}$ is the accuracy criterion that was determined through combining self-rating and acquaintance-rating for target j on item k. $Norm_k$ is an estimate of the average personality profile for item k on the judgment measure. Specifically, this normative estimate was gleaned by averaging the criterion scores on item k of each judgment measure across a large library of targets. It is important to note that prior to analysis, $Norm_k$ was subtracted from $TCrit_{jk}$. Doing so adjusted $TCrit_{jk}$ to match the operational definition of distinctive accuracy—how accurately judge i differentiates target j from the normative person, which is captured in the predictor $Norm_k$. After this adjustment, both predictors in the model ($TCrit_{jk}$ and $Norm_k$) were grand mean centred.

Given this data preparation, the intercept (β_{0ij}) of this level 1 model is the average predicted value of judge i's rating of target j on item k when $TCrit_{jk}$ and $Norm_k$ are at their mean levels. In a similar vein, β_{1ij} is the estimate of distinctive accuracy and represents the average change in judge i's rating of target j on item k for a one-unit increase to target j's criterion value on item k, while holding the normative estimate on item k at the mean value. Finally, β_{2ij} is the estimate of normativity and corresponds to the average change in judge i's rating of target j on item k for a one-unit increase in the normative profile on item k, while holding target j's criterion on item k at the mean value.

As seen at level 2, each of the level 1 predictors was modelled as a random effect. Specifically, the coefficients γ_{00}, γ_{10}, and γ_{20} represent the average intercept, distinctive accuracy, and normativity, respectively. Additionally, residual terms are included in each level 2 equation, u_{0i}, u_{1i}, and u_{2i}, to represent the residual variance attributed to the judge for the intercept, distinctive accuracy, and normativity, respectively; while u_{0j}, u_{1j}, and u_{2j} represent the residual variance attributed to the target for the intercept, distinctive accuracy, and normativity, respectively.[8]

In order to test information quantity in a continuous manner, video length (TIME) was entered as a moderator at level 2 of the base model [Equation (2)]. TIME was measured in minutes and was treated as a ratio variable such that a value of 0 represents 0 minute. Therefore, 0 TIME represents the extrapolation of accuracy coefficients for a video of 0 minute.

$$\begin{aligned} \beta_{0ij} &= \gamma_{00} + \gamma_{01}\mathrm{TIME} + u_{0i} + u_{0j} \\ \beta_{1ij} &= \gamma_{10} + \gamma_{11}\mathrm{TIME} + u_{1i} + u_{1j} \\ \beta_{2ij} &= \gamma_{20} + \gamma_{21}\mathrm{TIME} + u_{2i} + u_{2j} \end{aligned} \tag{2}$$

Video length is a ratio variable, even though only four lengths were used in the current study, and therefore, the continuous nature of this variable is accurately reflected in the TIME moderator. Here, the coefficients γ_{11} and γ_{21} represent the change in distinctive accuracy and normativity, respectively, for a 1-minute change in the video length.

To test the non-linearity of video length, a quadratic effect of time was tested.[9] This was carried out by adding the squared exposure time as a secondary moderator at level 2 [Equation (3)].

$$\begin{aligned} \beta_{0ij} &= \gamma_{00} + \gamma_{01}\mathrm{TIME} + \gamma_{02}\mathrm{TIME}^2 + u_{0i} + u_{0j} \\ \beta_{1ij} &= \gamma_{10} + \gamma_{11}\mathrm{TIME} + \gamma_{12}\mathrm{TIME}^2 + u_{1i} + u_{1j} \\ \beta_{2ij} &= \gamma_{20} + \gamma_{21}\mathrm{TIME} + \gamma_{22}\mathrm{TIME}^2 + u_{2i} + u_{2j} \end{aligned} \tag{3}$$

Now, the coefficients γ_{12} and γ_{22} represent the quadratic effect of time for distinctive and normativity, respectively.

[6]Despite this explicit effort, completion rates differed significantly across video length conditions ($\chi^2(3) = 10.36$, $p = .016$). Therefore, we explored individual difference characteristics of participants across the four conditions. No differences were observed for the variables of age ($F(3, 415) = 0.68$, $p = .57$), sex ($\chi^2(3) = 0.95$, $p = .81$), ethnic dominance coded as White/Caucasian versus other ($\chi^2(3) = 1.14$, $p = .77$), self-reported Big Five personality characteristics (all Fs(3, 421 [420 for Neuroticism]) < 1.14, $p > .33$), empathic response tendencies measured by the Davis (1983) Interpersonal Reactivity Index (all Fs(3, 423) < 1.70, $p > .17$), or satisfaction with life ($F(3, 423) = 2.13$, $p = .10$). Given the similarity across groups on these variables, the differences across groups in completion rates are unlikely to affect the interpretation of the results.

[7]Participants were also instructed to fill out other-report versions of the Ego-Control and Ego-Resiliency Scale (Block & Kremen, 1996), Subjective Happiness Scale (Lyubomirsky & Lepper, 1999), Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985), the dominance subscale of the Interpersonal Adjective Scale (Wiggins, Trapnell, & Phillips, 1988), and a measure of state affect, as well as self-report versions of the Interpersonal Reactivity Index (Davis, 1980), Satisfaction with Life Scale, BFI, Ego-Control and Ego-Resiliency Scale, and the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988). These additional measures were not included in the current analyses.

[8]It is standard when using the SAM (Biesanz, 2010) to estimate the dyadic effects. We attempted to fit all models with these effects included; however, some models did not converge. Thus, to keep all models consistent, we did not model the dyadic random effects for any of the analyses.

[9]We would like to thank a reviewer for the suggestion of testing the quadratic effect of time.

DOI: 10.1002/per

Table 2. Social accuracy model parameter estimates for all conditions combined and single conditions

	Condition				
	All (n = 431)	30 seconds (n = 103)	1 minute (n = 102)	3 minutes (n = 120)	5 minutes (n = 106)
Fixed effects (*SE*)					
Distinctive accuracy (γ_{10})	.176 (0.040)***	.152 (0.037)***	.164 (0.040)***	.198 (0.047)***	.186 (0.045)***
Normativity (γ_{20})	.485 (0.119)***	.412 (0.106)***	.478 (0.111)***	.534 (0.132)***	.511 (0.159)**
Judge random effects					
$\tau_{Distinctive}$	.117***	.115***	.127***	.115***	.107***
$\tau_{Normative}$	.387***	.347***	.426***	.395***	.370***
Target random effects					
$\tau_{Distinctive}$	.111***	.097***	.105***	.129***	.123***
$\tau_{Normative}$	.331***	.282***	.288***	.359***	.438***

Note: Standard errors of the fixed effects are in parentheses. Significance of the random effects was tested using nested chi-squared difference tests (Hox, 2010, pp. 47–50). *SE*, standard error. $*p < .05$. $**p < .01$. $***p < .001$.

RESULTS[10]

As shown in Table 2, significant levels of distinctive accuracy and normativity were achieved across all information quantity conditions combined and in each separate condition.[11] Additionally, there was significant variability in the distinctive accuracy and normativity scores across judges and targets (see judge and target random effects estimates in Table 2). It is noteworthy that normativity and distinctive accuracy varied approximately as much between judges as between targets. This is contrary to previous research (e.g. Biesanz, 2010), which has consistently found variance between targets to substantially exceed the variance between judges. Variance within judges is a necessary condition when exploring differences between judges. However, in the current study, random effect variance on the part of the judge (especially within randomly assigned groups) may actually lead to an underestimation of the effect of information quantity on perceiver accuracy.

The hypotheses were examined by adding video length as a moderator to the SAM equations. Results for all moderations can be found in Table 3. Video length was not a significant moderator for either distinctive accuracy or normativity. Exploring these data visually, when estimates of accuracy across traits were plotted by observation length (Figures 1a and 2a), a decrease in accuracy was observed in the 5-minute condition compared with the 3-minute condition. Therefore, a non-linear relationship was also tested by adding a squared length-of-observation variable as an additional moderator at level 2 of the SAM equations. This analysis revealed a significant non-linear relationship for distinctive accuracy but not for normativity. Although there was a change in the pattern of results by adding the quadratic terms, a nested chi-squared difference test (Hox, 2010, pp. 47–50) indicated that there was no significant level of improvement to model fit ($\chi^2(3) = 7.40$, $p = .06$),[12] and therefore, this quadratic effect will not be discussed further.

As an alternative method of calculating accuracy scores and testing the hypotheses, profile correlations were also calculated and compared across conditions with one-way analysis of variance and contrast analyses (see Data S3 for a detailed description of the analyses and results). This additional analysis makes it possible to more directly compare the results of the current study to the previous literature.[13] However, the reliability for distinctive accuracy correlations for individual judges across targets is quite low ($\alpha = .28$), and therefore, results should be interpreted with caution.[14] Consistent with the results of the SAM analyses, differences across the conditions did not reach statistical significance for normativity, $F(3, 426) = 1.83$, $p = .14$, $\eta^2 = .01$, and were just shy of the conventional cut-off for distinctive accuracy, $F(3, 426) = 2.51$, $p = .06$, $\eta^2 = .02$ (Figure S3.2). However, for distinctive accuracy and normativity, the data were consistent with the predicted increase over observation lengths based on contrast analyses (Figures S3.1 and S3.2).

ADDITIONAL ANALYSES BY TRAIT

After the initial analyses, exploratory analyses were conducted to examine whether the results differed for individual personality traits. To answer this question, five additional sets of analyses were used in which only the items for one trait were included in each model. This approach is consistent with previous research that has examined accuracy for specific traits in this manner using SAM (Biesanz & West,

[10]Additional analyses can be found in the supporting information on this study's OSF page. Data S1 includes one-way analysis of variance (and Tukey post hoc tests as appropriate) comparing means of distinctive accuracy and normativity across information quantities for traits overall and individual traits, as well as a contrast analysis for traits overall. Data S2 contains the main analyses from SAM, calculated using trait scores rather than individual items. Data S3 contains results using a profile correlation approach using the distinctive and normative profiles. Results based on these analyses are similar to the reported results.

[11]Multilevel models were estimated using the lme4 package (version 1.1–13) for R (version 3.4.1) with the lmerTest package (version 2.0–33) attached so that *df*s via Satterthwaite's approximation, and *p* values for fixed effects were available in summary output. This analysis can be found on this project's OSF page at https://osf.io/mrt8f/.

[12]See Data S2 for a description of results based on using trait scores instead of individual items. Results are similar, but there are some differences.

[13]Item-level or trait accuracy was not calculated due to only having eight targets per judge, which would result in unstable correlations.

[14]A main advantage of the SAM is that it is able to deal with this unreliability and estimate more stable accuracy and normativity scores.

DOI: 10.1002/per

Table 3. Time as a moderator of distinctive accuracy and normativity

	Linear effect			Non-linear effect		
Distinctive accuracy	γ_{11}	*SE*	*p* value	γ_{12}	*SE*	*p* value
All traits combined	0.006	0.004	.08	0.007	0.003	.008
Openness to Experience	−0.01	0.004	.006	−0.0006	0.003	.87
Conscientiousness	0.002	0.005	.68	−0.01	0.004	.008
Extraversion	0.03	0.007	<.001	0.02	0.006	.009
Agreeableness	−0.006	0.006	.29	−0.008	0.005	.10
Neuroticism	−0.002	0.005	.71	0.0008	0.004	.85
	Linear effect			Non-linear effect		
Normativity	γ_{21}	*SE*	*p* value	γ_{22}	*SE*	*p* value
All traits combined	0.02	0.01	.07	−0.01	0.009	.15
Openness to Experience	−0.02	0.01	.23	−0.006	0.01	.57
Conscientiousness	0.04	0.01	.01	−0.02	0.01	.05
Extraversion	−0.02	0.01	.14	0.0005	0.01	.96
Agreeableness	0.04	0.01	.001	−0.02	0.01	.13
Neuroticism	0.02	0.01	.09	−0.02	0.01	.11

Note: *SE*, standard error.

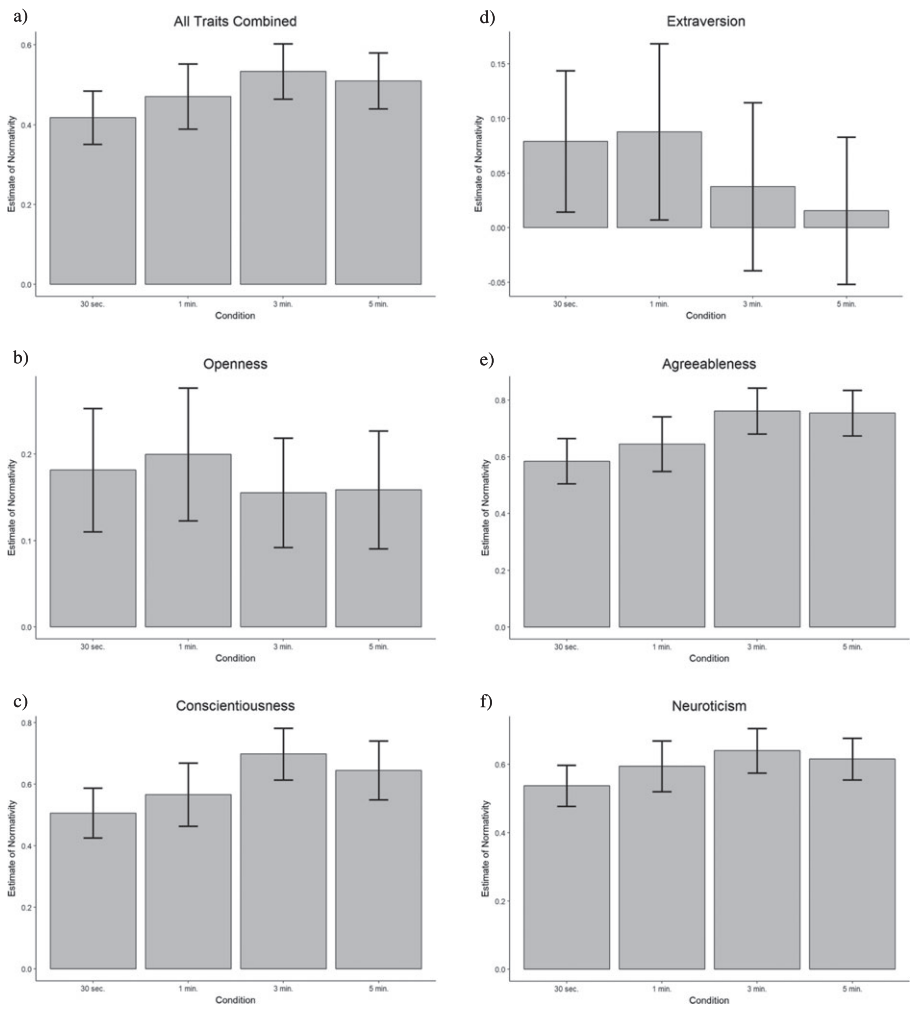

Figure 1. Normativity as a function of video observation length for all traits combined and for each trait separately. Error bars represent 95% confidence intervals. Take special note of the scaling on the *y*-axis as they are not identical across plots.

DOI: 10.1002/per

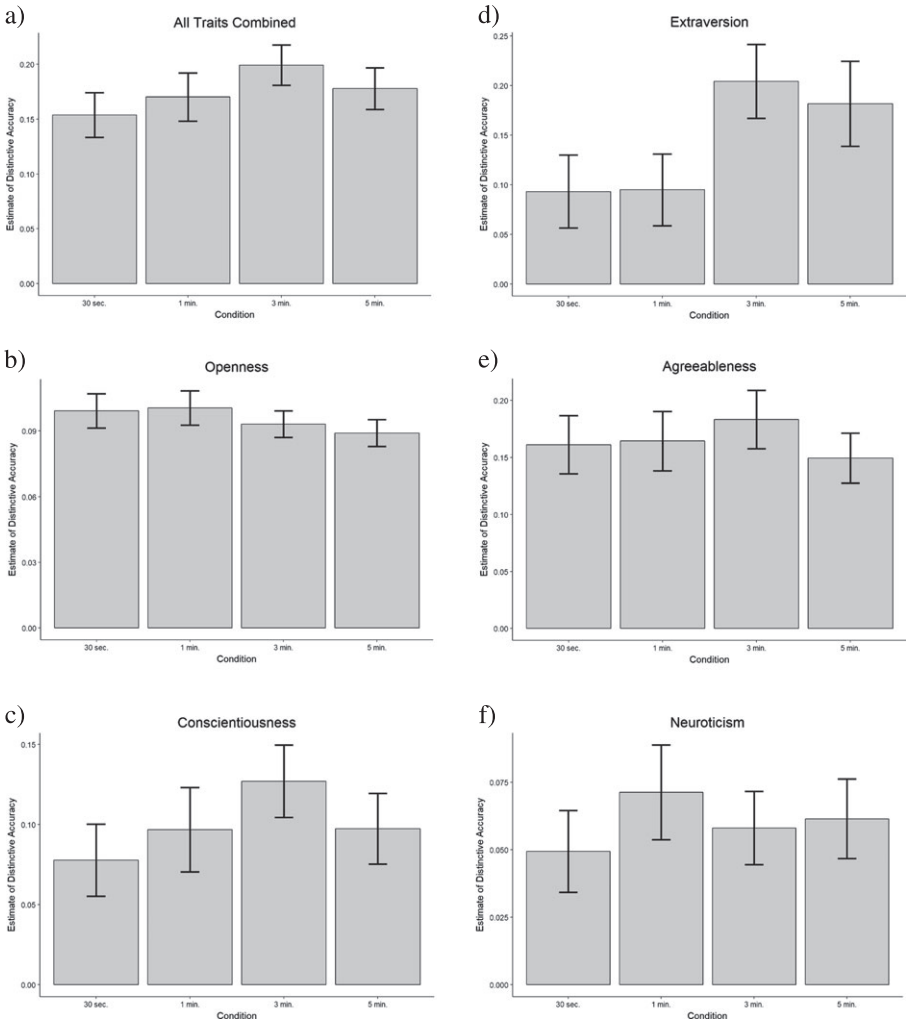

Figure 2. Distinctive accuracy as a function of video observation length for all traits combined and for each trait separately. Error bars represent 95% confidence intervals. Take special note of the scaling on the *y*-axis as they are not identical across plots.

2000; Letzring & Human, 2014) and with computing profile correlations across only the items for a given trait or averaging item-level correlations across items for a given trait (Borkenau, Brecke, Möttig, & Paelecke, 2009; Borkenau & Liebler, 1992; Gosling, Ko, Mannarelli, & Morris, 2002). These analyses allowed for an examination of whether the effect of information quantity on accuracy differed across personality traits. Due to the exploratory nature of these analyses, a priori hypotheses were not made, although it makes sense that more information would be especially useful for making distinctively accurate judgments of less visible traits for which it would be difficult to have access to enough cues within 30 seconds on which to base personality judgments. Less visible traits typically include Neuroticism and Openness to Experience (Funder & Dobroth, 1987; John & Robins, 1993). The analytic procedure for each trait paralleled that which was used previously. Specifically, significance of distinctive accuracy and normativity across all observation lengths was examined using the base model. Table 4 displays the fixed and random effect estimates for these trait-level models. Next, for each trait, video length (TIME) was entered as a continuous moderator of both distinctive accuracy and normativity, followed by the expanded model in which the quadratic effects were included.

Openness to Experience

The base model, across all experimental groups, indicated neither normativity nor distinctive accuracy for Openness was statistically significant (Table 4). That said, video length was a significant moderator for distinctive accuracy but not for normativity (Figures 1b and 2b). It is important to note that the estimate for the time moderator for distinctive accuracy was negative, which means that accuracy decreased slightly as time increased. The quadratic effects were not significant for distinctive accuracy or normativity, and the addition of the quadratic term did not improve model fit ($\chi^2(3) = 1.90$, $p = .59$).

Conscientiousness

Across experimental groups, a significant level of normativity for Conscientiousness was achieved but not distinctive accuracy. Video length was a significant moderator for normativity but not for distinctive accuracy (Figures 1c and 2c). There was also a significant non-linear effect of video length on both distinctive accuracy and normativity,

DOI: 10.1002/per

Table 4. Social accuracy model parameter estimates by individual trait across conditions

	Big Five personality trait				
	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Fixed effects (*SE*)					
Distinctive accuracy (γ_{10})	.095 (0.050)	.101 (0.058)	.146 (0.057)**	.165 (0.032)***	.060 (0.096)
Normativity (γ_{20})	.173 (0.095)	.608 (0.116)***	.054 (0.513)	.690 (0.099)***	.599 (0.096)***
Judge random effects					
$\tau_{Distinctive}$	.072***	.150***	.236***	.166***	.115***
$\tau_{Normative}$	.407***	.500***	.429***	.464***	.391***
Target random effects					
$\tau_{Distinctive}$	.138***	.161***	.151***	.083***	.270***
$\tau_{Normative}$	.260***	.319***	1.447***	.272***	.263***

Note: Standard errors of the fixed effects are in parentheses. Significance of random effects was tested using nested chi-squared difference tests (Hox, 2010, pp. 47–50). *SE*, standard error. *$p < .05$. **$p < .01$. ***$p < .001$.

and the addition of the quadratic term significantly improved model fit ($\chi^2(3) = 8.43$, $p = .04$). The quadratic effect was negative, which indicates that the initial relationship diminishes with increased observation lengths.

Extraversion

A significant level of distinctive accuracy, but not normativity, was found for judgments of Extraversion across conditions. Moreover, distinctive accuracy was significantly moderated by video length, but normativity was not (Figures 1d and 2d). The quadratic effect was also significant for distinctive accuracy and not for normativity. However, the addition of the quadratic term did not significantly improve model fit ($\chi^2(3) = 6.90$, $p = .08$) and is therefore not discussed further.

Agreeableness

In contrast to what was found for Openness, Conscientiousness, and Extraversion, significant levels were achieved for both normativity and distinctive accuracy for judgments of Agreeableness. Video length was a significant moderator of normativity but not distinctive accuracy (Figures 1e and 2e). The non-linear relations were not significant for normativity or distinctive accuracy, and inclusion of the quadratic term did not significantly improve model fit ($\chi^2(3) = 4.69$, $p = .20$).

Neuroticism

Similar to what was found for judgments of Conscientiousness, significant levels of accuracy were achieved for normativity but not distinctive accuracy across experimental conditions for judgments of Neuroticism. Neither distinctive accuracy nor normativity was significantly moderated by video length (Figures 1f and 2f). Likewise, the non-linear relations were not significant for distinctive accuracy or normativity, and inclusion of the quadratic term did not significantly improve model fit ($\chi^2(3) = 2.79$, $p = .43$).

Profile Correlations

As with the main analyses, the data for separate traits were also examined with the more traditional profile correlation approach (Data S3). Caution should be used in interpreting these results, as each judge–target correlation is only based on the 8–10 items for each trait, and reliabilities of the distinctive accuracy (range = −0.05–0.29) and normativity scores (range = 0.22–0.68) for individual judges across targets are low (Table S3.2). Distinctive accuracy differed significantly across groups for Extraversion, $F(3, 426) = 18.87$, $p < .001$, $\eta^2 = .12$, and fit a linear increase. Distinctive accuracy was higher at 3 minutes and 5 minutes than at 30 seconds and 1 minute but did not differ for 3 minutes versus 5 minutes or 30 seconds versus 1 minute (Table S3.3 and Figure S3.2). Normativity differed significantly across groups for Conscientiousness, $F(3, 426) = 3.63$, $p = .01$, $\eta^2 = .02$, and Agreeableness, $F(3, 422) = 3.51$, $p = .02$, $\eta^2 = .02$, although the only groups that differed significantly were 3 minutes and 30 seconds. Both traits fit a linear increase. These results are consistent with what was found using SAM.

DISCUSSION

It was predicted that both distinctive accuracy and normativity would increase as observation length increased, based on previous research that supports the acquaintanceship effect and the information quantity moderator. For distinctive accuracy when all traits were combined, the hypothesis was not supported: time was not a significant moderator. In addition, there was no significant change in the model fit for distinctive accuracy when the quadratic effect of time was added, indicating that a drop-off in distinctive accuracy at higher levels of information quantity did not better explain the pattern of results. For normativity with all traits combined, the hypothesis was also not supported, as both the linear and quadratic effects of time were not significant moderators. These findings are in contrast with previous results that found increases in levels of self-other agreement and overall accuracy as information quantity increased (Beer & Watson, 2010; Borkenau & Liebler, 1992). Despite this, previous research has also suggested that the effects of time on accuracy may be trait specific, and thus, examining traits overall does not necessarily provide a full understanding of this relationship (Brown & Bernieri, 2017).

When all traits were combined, both distinctive accuracy and normativity reached statistical significance for all

DOI: 10.1002/per

observation lengths. This finding is especially important in this context because it provides additional evidence that with access to only relatively small amounts of information, judges were able to describe targets in ways that reflected what the targets were really like and in ways that were consistent with what the average person is like, at levels that are greater than chance. These findings are consistent with previous research that has found statistically significant levels of accuracy for thin-slice exposures of less than 5 minutes (Ambady & Rosenthal, 1992; Beer & Watson, 2010; Borkenau & Liebler, 1992).

Additional analyses focused on each trait separately. For Openness, distinctive accuracy and normativity did not reach significant levels when all observation lengths were combined, indicating that there may not be enough relevant Openness cues within these videos for judges to achieve significant levels of distinctive accuracy and concurrently that judges are not relying on the knowledge of the average person when making judgments of this trait. Despite the fact that accuracy of Openness did not reach significance, there was a significant negative linear trend of distinctive accuracy across time. This may indicate that conflicting cues about Openness were being made available with longer observations, and therefore, distinctive accuracy decreased with more information; or perhaps, judges had a difficult time with correctly using more cues to judge Openness.

For Conscientiousness, a significant level of normativity was found when all conditions were combined, indicating that the knowledge of the average person is used in early judgments of Conscientiousness. Time significantly moderated normativity of Conscientiousness, and addition of the quadratic effect significantly improved model fit for both distinctive accuracy and normativity of this trait. This suggests that while there is a significant linear relationship between time and accuracy, this relationship diminishes in magnitude with increased exposure. This finding could indicate some level of unpredictability in accuracy of Conscientiousness with shorter quantities of information, as well as a possible drop in attention as time went on. It is possible that with longer quantities of information (such as 25 minutes, 50 minutes, 3 hours, etc.), the anticipated linear trend in accuracy across time would become more apparent (Blackman & Funder, 1998; Letzring et al., 2006).

For Extraversion, when all observation lengths were examined together, judges achieved a significant level of distinctive accuracy but not of normativity. It is possible that targets provided a high number of relevant Extraversion cues in these short interactions, even in the first 30 seconds, and therefore, judges were more likely to base judgments on the unique aspects of the targets rather than on what the average person is like. This is consistent with previous research that has found that traits with more observable cues, such as Extraversion, are more easily judged (Funder & Dobroth, 1987; John & Robins, 1993). Time was a significant moderator of distinctive accuracy for this trait, indicating that distinctive accuracy increased across thin slices of observation. This is in line with previous research that investigated the role of information quantity for accuracy of Extraversion, which found that this trait showed the greatest increase in accuracy over short interactions (5-minute conversation compared with zero acquaintance) compared with any other trait (Brown & Bernieri, 2017).

For Agreeableness, significant levels of distinctive accuracy and normativity were found when all observation lengths were combined, indicating that judges use cues provided by the target as well as knowledge of the average person in judgments with short observations. In addition, time was a significant moderator of normativity, but not distinctive accuracy, indicating that judges used more information about what the average person is like for judgments of Agreeableness as time increased. These results fit with previous research that has found a linear trend in accuracy of Agreeableness over time with short information quantities (Brown & Bernieri, 2017; Carney et al., 2007). Because a significant linear trend for distinctive accuracy of Agreeableness was not found, longer observation lengths may be required for judges to gather the cues required for more distinctively accurate judgments of this trait.

For Neuroticism, judgments across observation lengths achieved significance for normativity, which indicates that judges use information about what the average person is like when limited information about Neuroticism is available. In addition, time was not a significant moderator of either distinctive accuracy or normativity of Neuroticism, which fits with previous research indicating that accuracy of less visible traits requires greater quantities of information compared with other traits (Blackman & Funder, 1998; Letzring et al., 2006). Therefore, a longer acquaintanceship may be necessary for significant distinctive judgments of Neuroticism to take place and for accuracy of this trait to see noticeable improvements.

Implications and future directions

These outcomes are an important addition to the information quantity literature because previous work regarding information quantity has not always focused on judgments of personality traits and has also not examined how the components of distinctive accuracy and normativity are affected by short-length (e.g. thin slice) observations. While profile correlations have been a common approach in accuracy research when examining how well judges can rank traits within a target (or items within a trait), this method of analysis suffers from low power and instability of correlations (Schönbrodt & Perugini, 2013), because correlations are often across a small number of items for each judge–target pair. In contrast, the current research utilized the multilevel model SAM, which provided the ability to investigate multiple components of accuracy through a more powerful analytic approach. The findings indicate that even brief amounts of exposure to others via videotaped interactions or interviews can result in levels of distinctive accuracy and normativity that exceed chance levels but that increases in distinctive accuracy and normativity as a result of increased observation times of 5 minutes or less are only seen for some traits.

Based on these findings, it is clear that there are a number of differences in the effects of information quantity based on the trait of interest. For Conscientiousness and Extraversion,

DOI: 10.1002/per

the linear increase in accuracy decreased in magnitude across time, while for other traits (such as Openness and Agreeableness), accuracy increased or decreased in a linear fashion across short information quantities. In addition, the specific ways in which judges are accurate in terms of normativity and distinctive accuracy vary depending on the trait. These differences at the trait level are notable in that they indicate that certain traits benefit more from increases in information quantity when time is limited compared with others and that judges achieve accuracy in different ways depending on the trait. These findings are in alignment with previous research that has found differences in how length of acquaintanceship affects accuracy depending on the trait and point to the idea that examination of separate traits is essential to more fully understand the effects of information quantity on accuracy (Brown & Bernieri, 2017; Carney et al., 2007).

It is interesting to note that across all conditions, judges achieved significant levels of distinctive accuracy for Extraversion and Agreeableness but not for the other traits. It is interesting that this finding applies to the trait of Agreeableness, because this is a trait that has been found to be judged with lower levels of accuracy compared with other traits at short information quantities (Carney et al., 2007; Human & Biesanz, 2011). This indicates that relevant cues for these traits are being made visible in the short interactions used in the current stimulus materials. In contrast, the traits of Conscientiousness and Neuroticism were judged with significant levels of normativity, which indicates that judgments of these traits are more in line with what the average person is like, and that possibly less relevant information was visible which necessitated judgments that were more prototypical and less distinctive. Agreeableness was judged with significant levels of both normativity and distinctive accuracy, which indicates that judges use a combination of visible target cues and knowledge of the average person in judgments of Agreeableness when only limited information is available. Openness to Experience was not accurately judged in either regard, which could mean that neither type of information is readily available in such a short time frame, and that a longer acquaintanceship is necessary to achieve accuracy for this trait. With shorter quantities of information, there appear to be a number of subtleties present in how judges perceive targets, in that cues provided by a target are more likely to be used in judgments of some traits, while knowledge of the average person is more likely to be used in judgments of other traits. It is unknown if these findings would generalize to brief in-person interactions, and therefore, future research should examine relatively short face-to-face interactions of differing lengths to determine how this influences distinctive accuracy and normativity of personality judgment.

While examination of separate traits provided a greater understanding of the role of variations in information quantity across relatively short amounts of time on different types of accuracy, there are still several unanswered questions about factors of the judge, target, and situation that could influence this effect. The current study did not examine how aspects of the judge influence this relationship, in that certain judges may benefit more from smaller amounts of information compared with others, and this difference could also depend on the trait of interest. In addition, the types of situations and contexts that are used within the stimulus materials may play a role in the effects of information quantity on accuracy, especially when investigating thin slices of information. Future research should also investigate cue type and frequency to more fully understand the amount of information judges are being exposed to in ways other than measurements of time. Due to the extensive nature of the behavioural coding required for this type of project (Letzring & Human, 2014) and the fact that only eight targets were used (two in each situation), this type of analysis was not within the scope of the current research but is a worthy avenue for future research.

A final recommendation for future research is to pool together large bodies of data from previous accuracy studies to conduct similar analyses as utilized in the current study. As previous accuracy research has utilized different statistical techniques (profile correlations, item-level correlations, SAM, etc.), with comparison being complex across different methods, it should be a priority for accuracy researchers to compare across work with more ease and test previous findings with more powerful approaches. It is likely that this would produce more discerning conclusions, as large amounts of data from multiple studies could be analysed using the same statistical methods and interpretations.

Limitations

One possible limitation of the current study is that judges were recruited using MTurk, which means that there was no control over the environment in which the observations and ratings were made (Crump, McDonnell, & Gureckis, 2013). Participants may have been distracted (Clifford & Jerit, 2014), and attention to the protocol (especially the stimulus videos) may have been less than ideal. This caveat may explain the decrease in accuracy in the 5-minute condition compared with the 3-minute condition that was best explained by a non-linear effect in a few of the analyses. It is possible that participants in the 5-minute video condition may have simply lost interest and did not pay attention to the entire video. However, the finding that both types of accuracy were above chance levels for all conditions when traits were combined is evidence that participants were attending to the videos and making ratings based on those videos on a level that we would expect from participants in a monitored lab setting, indicating that this type of sample is appropriate for accuracy research. Furthermore, there is some ecological validity to this methodological issue in that it is also likely for people to become distracted in real-life interactions and not pay attention to the people whom they are observing or with whom they are interacting. The potential problems with an MTurk sample were also minimized by requiring participants to pass attention checks and complete most of the study for their data to be included in analyses (Hauser & Schwarz, 2016). Despite the possible limitations of MTurk research, the findings demonstrate sufficient attention was paid to the instructions, videos, and questionnaires/measures.

DOI: 10.1002/per

Another possible limitation comes from the stimulus materials used for this study. While previous research has utilized videos as a way of assessing accuracy (e.g. Biesanz & Human, 2010; Colman et al., 2017; Letzring, 2015), it does lack ecological validity in that participants were not interacting with targets face-to-face. Despite this limitation, using videos instead of face-to-face interactions allowed for the creation of conditions in which all participants in a single condition were exposed to the same stimulus, and participants in different conditions were all exposed to the same targets. This, in addition to random assignment, allowed for a causal test of the link between observation length and judgmental accuracy, which is uncommon in research on judgmental accuracy and therefore a strength of this particular study. Future studies could also include situations where the targets interact with people that they know and/or are performing tasks alone in order to broaden the generalizability of the situations. Targets could also be observed engaging in situations or tasks of their own choosing as opposed to following an experimenter's specific instructions.

To increase generalizability of the findings, five different situations were presented in the videos. All of the situations took place in the lab and were of a medium situational strength (Marshall & Brown, 2006; Snyder & Ickes, 1985), meaning that behaviour was allowed to vary somewhat, depending on the personality of the targets. In addition, each of the situations took place between two or three people who had not met before. Therefore, the findings can at least be generalized to other contexts in which strangers interact in situations where there are only moderate expectations for appropriate behaviour, and therefore, behaviour is likely to provide some relevant cues to personality. These results may also generalize to observations of targets who know each other, and to weaker or stronger situations, but those are questions for future research. Certainly, there is a relation between the type of context in which targets are observed and the effect of information quantity on accuracy. For example, strong situations with very little variation in behaviour across targets may require especially high quantities of information for accuracy to increase, whereas weak situations may require much smaller quantities for the same amount of increase in accuracy. Also, cues relevant to some traits will be more easily available in some situations than others, and the traits related to these relevant cues will benefit more from additional observation time than other traits.

Another possible limitation of the current study was the use of only eight targets across all judges. Previous accuracy research has typically made a trade-off between the numbers of judges and targets, depending on the focus of the research, and it has been common to see smaller numbers of targets when a large sample of judges is needed (Funder & Colvin, 1988 used eight targets; Letzring, 2015 used 10 targets). For example, Biesanz et al. (2007) used a design in which participants were rated by up to two other acquaintances, without the use of any stranger ratings. Research in information quantity focusing on the target has utilized 150 targets with only two judges per target (Funder & Colvin, 1988), as well as 100 targets with only 24 judges (Borkenau & Liebler, 1992). Work focusing on the judge has utilized 334 judges and 30 targets (Carney et al., 2007). Other work has used even smaller numbers of judges and targets (176 judges who each rated five to seven targets in Brown & Bernieri, 2017; 180 judges who each rated three targets in Letzring et al., 2006). Within the information quantity literature, there has been little consensus as to the most ideal number of targets and judges, with obvious variations across studies depending on the question of interest.

Based on the wide variety in target sample size throughout previous research, it is difficult to identify the ideal ratio of targets to judges, especially in information quantity research that focuses on the ability of the judge. Despite this lack of consensus, the use of eight targets is a limitation that makes generalization to other target pools difficult, and it is possible that research investigating a different set of targets may find slightly different outcomes. Despite this limitation, the targets selected for this study were chosen in an effort to represent variability across traits, affect, and situations to increase the generalizability of the findings. Future research should aim to utilize a larger pool of targets to more fully understand the nuances present within accuracy at short information quantities, especially considering there are still many unanswered questions regarding the effects of target-specific factors in understanding the role of information quantity on accuracy.

CONCLUSION

Overall, videos of short lengths all yielded significant levels of distinctive accuracy and normativity when traits were combined. In addition, a linear trend in accuracy as information quantity increased was not found using a more powerful analytic approach compared with previous research, indicating that differences in information quantity at small levels may have limited effects on accuracy of combined judgments of the Big Five traits. When traits were examined individually, different results emerged depending on the trait of interest. The findings have implications for methodological designs within the field of personality judgment accuracy that aim to use video target stimuli and MTurk (or other online crowdsourcing platforms) and indicate that even short observation lengths can result in accurate judgments. This study also advocates for the use of SAM in analysing accuracy, as this approach is more statistically powerful than traditional profile and item-level correlations. Finally, this research demonstrates the importance of investigating the separate components of accuracy rather than just accuracy as a single construct, as well as the benefit of examining traits separately in an attempt to more fully understand and explain the reasons for achieving more accurate judgments.

ACKNOWLEDGEMENT

Data collection was supported by an Idaho State University Office for Research Faculty Seed Grant (5-2014) to Tera D. Letzring.

DOI: 10.1002/per

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. ANOVA Results
Table S1.1 Comparing means of distinctive accuracy and normativity across levels of information quantity with all traits combined.
Table S1.2 Differences between distinctive accuracy for levels of information quantity with all traits combined.
Table S1.3 Contrast analyses testing the prediction that distinctive accuracy and normativity would increase as observation length increased with either equidistant contrast weights (-3, -1, +1, +3) or unequal-distant contrast weights that represent the actual differences in time between groups (-15, -11, +5, +21).
Table S1.4 Comparing means of distinctive accuracy or normativity across levels of information quantity for individual traits.
Table S1.5 Tukey post-hoc tests to test for significant differences between means of either distinctive accuracy or normativity for levels of information quantity for individual traits.
Data S2. Main Analyses Using Trait Scores
Table S2.1 Social Accuracy Model Parameter Estimates for All Conditions Combined and Single Conditions, Using Trait Scores
Figure S2.1. Normativity and distinctive accuracy as a function of video observation length for all traits combined, when using trait scores in the SAM analysis. Error bars represent 95% confidence intervals. Take special note of the scaling on the y-axes, as they are not identical across plots.
Data S3. Profile Correlations
Table S3.1 Profile Distinctive Accuracy Scores for All Conditions Combined and Single Conditions
Table S3.2 Profile Scores for Individual Traits for All Conditions Combined and Tests of Differences Across Conditions
Table S3.3 Results of Tukey Post-Hoc Tests
Figure S3.1. Normativity calculated with profile correlations as a function of video observation length for all traits combined and for each trait separately. Error bars represent 95% confidence intervals. Take special note of the scaling on the y-axis as they are not identical across plots.
Figure S3.2. Distinctive accuracy as calculated with profile correlations as a function of video observation length for all traits combined and for each trait separately. Error bars represent 95% confidence intervals. Take special note of the scaling on the y-axis as they are not identical across plots.
Data S4. Constraints on Generality (COG) Statement (Simons, Shouda, & Lindsay, 2017).

REFERENCES

Allik, J., Borkenau, P., Hrebícková, M., Kuppens, P., & Realo, A. (2015). How are personality trait and profile agreement related? *Frontiers in Psychology*, *6*, 1–11 https://doi.org/10.3389/fpsyg.2015.00785.

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

Beer, A. (in press). Information as a moderator of accuracy in personality judgment. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment*. New York, NY: Oxford University Press.

Beer, A., & Watson, D. (2010). The effects of information and exposure on self-other agreement. *Journal of Research in Personality*, *44*, 38–45 https://doi.org/10.1016/j.jrp.2009.10.002.

Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, *45*, 853–885 https://doi.org/10.1080/00273171.2010.519262.

Biesanz, J. C. (2018). Interpersonal perception models. In V. Zeigler-Hill, & T. K. Shackelford (Eds.), *The SAGE Handbook of Personality and Individual Differences Volume I: The Science of Personality and Individual Differences*. Thousand Oaks, CA, USA: Sage Publications. https://doi.org/10.4135/9781526451163.n24.

Biesanz, J. C. (2019). The social accuracy model. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment*. New York, NY: Oxford University Press Chapter in preparation.

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: Accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, *21*, 589–594 https://doi.org/10.1177/0956797610364121.

Biesanz, J. C., & West, S. G. (2000). Personality coherence: Moderating self-other profile agreement and profile consensus. *Journal of Personality and Social Psychology*, *79*, 425–437 https://doi.org/10.1037//0022-3514.79.3.425.

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*, 119–135 https://doi.org/10.1037/0022-3514.92.1.119.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology*, *34*, 164–181 https://doi.org/10.1006/jesp.1997.1347.

Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. *Journal of Personality and Social Psychology*, *70*, 349–361 https://doi.org/10.1037/0022-3514.70.2.349.

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality*, *43*, 703–706. https://doi.org/10.1016/j.jrp.2009.03.007.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 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*, 599–614 https://doi.org/10.1037/0022-3514.86.4.599.

Brown, J. A., & Bernieri, F. (2017). Trait perception accuracy and acquaintance within groups: Tracking accuracy development. *Personality and Social Psychology Bulletin*, *43*, 716–728 https://doi.org/10.1177/0146167217695557.

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*, 1054–1072 https://doi.org/10.1016/j.jrp.2007.01.004.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*, 120–131 https://doi.org/10.1017/xps.2014.5.

Colman, D. E. (in press). Characteristics of the judge that are related to accuracy of personality trait judgments. In T. D. Letzring, & J.

DOI: 10.1002/per

S. Spain (Eds.), *The Oxford handbook of accurate personality judgment*. New York, NY: Oxford University Press.

Colman, D. E., Letzring, T. D., & Biesanz, J. C. (2017). Seeing and feeling your way to accurate personality judgments: The moderating role of perceiver empathic tendencies. *Social Psychological and Personality Science*, *8*, 806–815. https://doi.org/10.1177/1948550617691097.

Colvin, C. R. (1993). "Judgable" people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, *64*, 861–873 https://doi.org/10.1037/0022-3514.64.5.861.

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

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

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, *52*, 177–193 https://doi.org/10.1037/h0040957.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*, e57410 https://doi.org/10.1371/journal.pone.0057410.

Davis, M. H. (1980). Individual differences in empathy: A multidimensional approach. *Dissertation Abstracts International*, *40*, 3480.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126 https://doi.org/10.1037/0022-3514.44.1.113.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75 https://doi.org/10.1207/s15327752jpa4901_13.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191 https://doi.org/10.3758/BF03193146.

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

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

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*, 149–158 https://doi.org/10.1037/0022-3514.55.1.149.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, *52*, 409–418 https://doi.org/10.1037/0022-3514.52.2.409.

Furr, R. M. (2008). Framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*, 1267–1316 https://doi.org/10.1111/j.1467-6494.2008.00521.x.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe, Vol. 7* (pp. 7–28). Tilburg, the Netherlands: Tilburg University Press.

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*, 379–398. https://doi.org/10.1037//0022-3514.82.3.379.

Hall, J. A., Back, M. D., Nestler, S., Frauendorfer, D., Schmid Mast, M., & Ruben, M. A. (2018). How do different ways of measuring individual difference in zero-acquaintance personality judgment accuracy correlate with each other? *Journal of Personality*, *86*, 220–232. https://doi.org/10.1111/jopy.12307.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407 https://doi.org/10.3758/s13428-015-0578-z.

Hollander, E. P. (1956). The friendship factor in peer nominations. *Personnel Psychology*, *9*, 435–447 https://doi.org/10.1111/j.1744-6570.1956.tb01078.x.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (G. A. Marcoulides Ed. 2nd ed.). New York, NY: Routledge. https://doi.org/10.4324/9780203852279.

Human, L. J., & Biesanz, J. C. (2011). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. *Journal of Personality and Social Psychology*, *101*, 202–216 https://doi.org/10.1037/a0023782.

Human, L. J., & Biesanz, J. C. (2013). Targeting the good target: An integrative review of the characteristics and consequences of being accurately perceived. *Personality and Social Psychology Review*, *17*, 248–272 https://doi.org/10.1177/1088868313495593.

Human, L. J., Jackson, J. J., & Biesanz, J. C. (2008, February). *How do impressions change over time? Length of exposure and changes in differential and stereotype accuracy*. Poster presented at the meeting of the Society for Personality and Social Psychology, Albuquerque, NM.

IPIP NEO-PI domains. (n.d.). *International personality item pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences (<http://ipip.ori.org/>)*. Internet Web Site.

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology*, *60*, 348–361. https://doi.org/10.1037/0022-3514.60.3.348.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York, NY, US: Guilford Press.

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*, 521–551 https://doi.org/10.1111/j.1467-6494.1993.tb00781.x.

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

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

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin*, *116*, 245–258 https://doi.org/10.1037/0033-2909.116.2.245.

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice*, *9*, 300–311 https://doi.org/10.1093/clipsy/9.3.300.

Kolar, D. W. (1996). *Individual differences in the ability to accurately judge the personality characteristics of others*. (Unpublished doctoral dissertation), University of California, Riverside, Riverside, CA.

Krzyzaniak, S. L., & Letzring, T. D. (in press). Characteristics of traits that are related to accuracy of personality judgments. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment*. New York, NY: Oxford University Press.

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*, 21–48 https://doi.org/10.1111/1467-6494.t01-1-00005.

DOI: 10.1002/per

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*, 668–682 https://doi.org/10.1037/a0018771.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, *42*, 914–932 https://doi.org/10.1016/j.jrp.2007.12.003.

Letzring, T. D. (2015). Observer judgmental accuracy of personality: Benefits related to being a good (normative) judge. *Journal of Research in Personality*, *54*, 51–60 https://doi.org/10.1016/j.jrp.2014.05.001.

Letzring, T. D., & Colman, D. E. (2018). *The Idaho Test of Accurate Person Perception (ITAPP): Preliminary results from the creation of a standardized measure of individual differences in the accuracy of perceiving others*. Manuscript under revision.

Letzring, T. D., & Funder, D. C. (2018). Interpersonal accuracy in trait judgments. In V. Zeigler-Hill & T. K. Shackelford (eds.), *The SAGE handbook of personality and individual difference (volume 3: Applications of personality and individual differences)*. Thousand Oaks, CA: Sage Publications, Inc. https://doi.org/10.4135/9781526451248.n11

Letzring, T. D., & Funder, D. C. (in press). The realistic accuracy model. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment*. New York, NY: Oxford University Press.

Letzring, T. D., & Human, L. J. (2014). An examination of information quality as a moderator of accurate personality judgment. *Journal of Personality*, *82*, 440–451 https://doi.org/10.1111/jopy.12075.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, *91*, 111–123 https://doi.org/10.1037/0022-3514.91.1.111.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*, 137–155 https://doi.org/10.1023/A:1006824100041.

Marshall, M. A., & Brown, J. D. (2006). Trait aggressiveness and situational provocation: A test of the Traits as Social Sensitivities (TASS) model. *Personality and Social Psychology Bulletin*, *32*, 1100–1113 https://doi.org/10.1177/0146167206288488.

Mignault, M., & Human, L. J. (in press). The good targets of personality judgments. In T. D. Letzring, & J. S. Spain (Eds.), *The Oxford handbook of accurate personality judgment*. New York, NY: Oxford University Press.

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

Park, B., Kraus, S., & Ryan, C. S. (1997). Longitudinal changes in consensus as a function of acquaintance and agreement in liking. *Journal of Personality and Social Psychology*, *72*, 604–616 https://doi.org/10.1037/0022-3514.72.3.604.

Paulhus, D. L., & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology*, *69*, 1233–1242 https://doi.org/10.1037/0022-3514.69.6.1233.

Rogers, K. H., & Biesanz, J. C. (2014). The accuracy and bias of interpersonal perceptions in intergroup interactions. *Social Psychological and Personality Science*, *5*, 918–926 https://doi.org/10.1177/1948550614537307.

Rogers, K. H., & Biesanz, J. C. (2015). Knowing versus liking: Separating normative knowledge from social desirability in first impressions of personality. *Journal of Personality and Social Psychology*, *109*, 1105–1116 https://doi.org/10.1037/a0039587.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612 https://doi.org/10.1016/j.jrp.2013.05.009.

Simmons, J., Nelson, L., & Simonsohn, U. (2012). A 21 word solution. *The Official Newsletter of the Society for Personality and Social Psychology*, *26*, 4–7 https://doi.org/10.2139/ssrn.2160588.

Simons, D. J., Shouda, Y., & Lindsay, S. D. (2017). Constraints on Generality (COG): a proposed addition to all emperical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128 https://doi.org/10.1177/1745691617708630.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey, & E. Aronson (Eds.), *Handbook of social psychology, Vol. II* (3rd ed., pp. 883–947). New York, NY: Random House.

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

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality and Social Psychology*, *71*, 267–295.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070 https://doi.org/10.1037/0022-3514.54.6.1063.

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, *23*, 517–530 https://doi.org/10.1207/s15327906mbr2304_8.

DOI: 10.1002/per