Interdisciplinary Public Policy Mainz

IPP

Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

In Your Eyes Only? Discrepancies and Agreement Between Self- and Other-Reports of Personality From Age 14 to 29

Julia M. Rohrer, Boris Egloff, Michal Kosinski,

David Stillwell und Stefan C. Schmukle

2017

Discussion paper number 1702

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
wiwi.uni-mainz.de

Contact details:

Julia M. Rohrer
University of Leipzig
Department of Psychology
04109 Leipzig, Germany

julia.rohrer@uni-leipzig.de

Boris Egloff
Johannes Gutenberg University of Mainz
Department of Psychology
55099 Mainz, Germany

Michal Kosinski
Stanford University
Stanford Graduate School of Business
Stanford, CA 94305

David Stillwell
University of Cambridge
Cambridge Judge Business School
Cambridge CB2 1AG, UK

Stefan C. Schmukle
University of Leipzig
Department of Psychology
04109 Leipzig, Germany

In Your Eyes Only? Discrepancies and Agreement Between Self- and Other-Reports of Personality From Age 14 to 29

Julia M. Rohrer[1], Boris Egloff[2], Michal Kosinski[3], David Stillwell[4], Stefan C. Schmukle[1]

[1]University of Leipzig

[2]University of Mainz

[3]Stanford University

[4]University of Cambridge

Journal of Personality and Social Psychology, in press.

Author Note

All analysis scripts and the final data sets can be retrieved via https://osf.io/x3sw2/. Raw data are available after registering as a collaborator on the myPersonality homepage, http://www.mypersonality.org.

Michal Kosinski and David Stillwell co-own the mypersonality.org database.

Correspondence concerning this article should be addressed to Julia M. Rohrer, Department of Psychology, University of Leipzig, Neumarkt 9-19, 04109 Leipzig, Germany. E-mail: julia.rohrer@uni-leipzig.de

Abstract

Do others perceive the personality changes that take place between the ages of 14 and 29 in a similar fashion as the aging person him- or herself? This cross-sectional study analyzed age trajectories in self- versus other-reported Big Five personality traits and in self-other agreement in a sample of more than 10,000 individuals from the *myPersonality Project*. Results for self-reported personality showed maturation effects (increases in extraversion, conscientiousness, openness to experience, and emotional stability), and this pattern was generally also reflected in other-reports, albeit with discrepancies regarding timing and magnitude. Age differences found for extraversion were similar between the self- and other-reports, but the increase found in self-reported conscientiousness was delayed in other-reports, and the curvilinear increase found in self-reported openness was slightly steeper in other-reports. Only emotional stability showed a distinct mismatch with an increase in self-reports, but no significant age effect in other-reports. Both the self- and other-reports of agreeableness showed no significant age trends. The trait correlations between the self- and other-reports increased with age for emotional stability, openness, agreeableness, and conscientiousness; by contrast, agreement regarding extraversion remained stable. The profile correlations confirmed increases in self-other agreement with age. We suggest that these gains in agreement are a further manifestation of maturation. Taken together, our analyses generally show commonalities but also some divergences in age-associated mean level changes between self- and other-reports of the Big Five, as well as an age trend towards increasing self-other agreement.

Word count: 239

Keywords: personality maturation, other-reports, self-other agreement, Big Five, personality cross-ratings

In Your Eyes Only? Discrepancies and Agreement Between Self- and Other-Reports of Personality From Age 14 to 29

Personality—the way in which we individually experience the world and engage in it—is not cast in stone. A large body of studies has shown that personality can change over the life course (Caspi, Roberts, & Shiner, 2005; McAdams & Olson, 2010; Roberts & DelVecchio, 2000; Roberts, Walton, & Viechtbauer, 2006), even well into advanced age (Kandler, Kornadt, Hagemeyer, & Neyer, 2015; Wagner, Ram, Smith, & Gerstorf, 2015). Such changes are especially pronounced during the transition from adolescence to young adulthood. Coming of age has been associated with normative personality changes toward greater maturity: For example, young people tend to become more socially confident, less angry (Roberts, Caspi, & Moffitt, 2001), more conscientious, and more agreeable (Srivastava, John, Gosling, & Potter, 2003). However, these findings have mostly been based on self-reports of personality, and there have been calls for more methodological diversity in research on personality change (Roberts & Mroczek, 2008). In the present study, we analyzed whether age effects in self-reported personality are also reflected in other-reported personality. Furthermore, we wanted to evaluate whether agreement between self- and other-reports is itself subject to age effects: Is maturation accompanied by higher congruence between one's self-view and the ratings of others?

Personality Maturation in Self-Reports

In the transition from late adolescence to young adulthood, people's self-reports of personality indicate that they become more mature. Expressed in terms of the Big Five personality traits, the current predominant taxonomy in personality psychology (John, Naumann, & Soto, 2008), people grow toward higher emotional stability, higher conscientiousness, and higher agreeableness (Roberts & Mroczek, 2008; Roberts et al., 2006; Soto, John, Gosling, & Potter, 2011; Srivastava et al., 2003).

The overall evidence for maturation in self-reported personality from adolescence to adulthood is persuasive. Positive age trends have been identified in large-scale cross-sectional (e.g., Soto et al., 2011), longitudinal (see Roberts et al., 2006, for a meta-analysis), and cross-cultural studies (Bleidorn et al., 2013). They have been linked to both genetic and environmental factors (see Bleidorn, 2015, for an overview) and are shaped by role transitions and life events (e.g., Bleidorn et al., 2013; Specht, Egloff, & Schmukle, 2011).

However, this picture becomes more complicated if we zoom in on various aspects of this general trend. The results on extraversion and openness have been less consistent. Considering extraversion, Roberts and Mroczek (2008) suggested that different subdimensions follow different age trends: Individuals tend to become more socially dominant, but social vitality increases only slightly in adolescence and then decreases again. Regarding openness to experience, Soto et al. (2011) suggested a curvilinear pattern that reconciles contradictory findings: During adolescence, openness first decreases but then increases again.

Furthermore, cross-sectional as well as longitudinal studies of personality development in younger samples from age 12 to 18 (McCrae et al., 2002; Allik, Laidra, Realo, & Pullmann, 2004; Pullmann, Raudsepp, & Allik, 2006) have found little evidence of maturation, and some have even found negative age trends that were partly moderated by gender. For example, McCrae et al. (2002) found that girls but not boys became *less* emotionally stable from age 12 to 18, potentially marking the emergence of the gender differences that have been observed in adults. Thus, the overall pattern of personality changes toward greater maturity needs to be qualified with respect to traits, age groups, and potential gender differences.

The Limitations of Self-Reports in Research on Maturation

The fact that the largest part of research on personality maturation relies solely on self-reports is no surprise. Self-reports are probably the most obvious measure of personality: Who else would have access to such a large quantity and breadth of information if not the individual him-/herself? Furthermore, self-reports are the most convenient and the most established way to measure personality. Alternative modes of assessment such as informant-reports or behavioral observations are more intricate and more expensive. Besides, the dominance of self-report measures is not limited to research on personality maturation or to personality psychology.

However, a reliance on self-assessment has been shown to be problematic across domains (Dunning, Heath, & Suls, 2004). More specifically, self-reports of personality are affected by response styles as they are susceptible to self-presentation and effects of social desirability (see e.g., Paulhus & Vazire, 2007). Such biases are dangerous for research on personality development: Age differences in these confounding variables could undermine the validity of the concept of maturation. Indeed, Soubelet and Salthouse (2011) found that older respondents scored higher on social desirability measures and that accounting for these differences attenuated positive age trends in personality. Thus, age effects in self-reports of personality call for independent validations that employ other sources of information, such as other-reports of personality.

Personality Maturation in Other-Reports

The literature on age effects in other-reported personality is comparatively sparse, but a number of cross-sectional and longitudinal studies have addressed this topic.

In the large-scale studies by McCrae and Terracciano (2005, $N > 10{,}000$) and Allik et al. (2009, $N > 7{,}000$), participants were randomly assigned to rate the personality of either a college-aged individual or an adult (older than 40 years or 50 years, respectively). McCrae

and Terracciano found that for the college-aged targets (18 to 21 years), a group that could be expected to show effects of personality maturation, other-reports of conscientiousness and openness increased with age, extraversion decreased slightly, and agreeableness and emotional stability showed no age trends. Allik et al.'s (2009) results for college-aged targets (17 to 23 years) reflected the same trends only in part: Again, conscientiousness increased, extraversion decreased, and no remarkable effects were found for emotional stability. However, in this study, agreeableness increased, and openness actually *decreased*.

De Bolle et al. (2015) used the same design in a cross-cultural sample (N = 4,850) but focused on younger age groups (targets between either 12 and 14 or 15 and 17 years) and analyzed gender differences in age trends. They found multiple interactions between gender and age in their cross-sectional analysis. For example, both boys and girls increased in emotional stability, but the trend was more pronounced in boys, which could mark the emergence of a gender difference in emotional stability from the observer's perspective.

Per design, all three studies captured only a very narrow range of ages, which limited their interpretability regarding late adolescence and early adulthood. Furthermore, they did not include targets' self-reports, which are crucial for comparing age trends in other-reports with the corresponding trends in self-reports.

Such a comparison was possible in a study by McCrae et al. (2004) in which both self- and other-reported personality was observed in a total of 1,505 individuals from the Czech Republic and Russia (McCrae et al., 2004). About 800 individuals fell within the age range relevant for personality maturation (15 to 29 years), and the sample was split into multiple age groups across the range of 15 to 60+ years of age. Self-reports showed mostly linear increases across the entire age range in agreeableness and conscientiousness, a u-shaped trend in emotional stability, and decreases in extraversion and openness. Age effects in other-reports appeared mitigated yet showed the same trends (for extraversion,

agreeableness, conscientiousness, and openness) or failed to reach statistical significance (for emotional stability).

A different perspective on other-reported maturation has been offered by longitudinal studies that compared self-reports with other-reports made by important others of the same individuals at various ages. Göllner et al. (2016) analyzed the personality development of nearly 3,000 students from age 11 to age 14. In addition to collecting self-reports, the authors asked parents to report on their children's personality. Because of the young age of the sample, self-reports did not follow the pattern of maturation that is typically found in older adolescents: Agreeableness, conscientiousness, and openness decreased, extraversion increased, and emotional stability remained stable. The parent-reports confirmed these trends only in part: Agreeableness and openness also decreased in the parent-reports, but in contrast to the self-reports, conscientiousness remained stable, extraversion decreased, and emotional stability increased. A similar longitudinal study (Luan, Hutteman, Denissen, Asendorpf, & van Aken, 2016) covered personality changes between the ages of 12 and 29 ($N = 186$) and 12 and 18 ($N = 574$), respectively. The self-reports in both studies indicated no changes in extraversion, no changes in emotional stability, and increases in agreeableness, conscientiousness, and openness to experience. Again, the parent-reports drew quite a different picture. In the first sample, parents perceived greater overall maturation in their children than the children perceived themselves. However, in the second sample, parents perceived *less* pronounced effects of maturation or even negative personality changes. It is interesting that the second sample also included ratings by siblings, which—in contrast to the parent-ratings—followed the same age trends as the self-reports for all of the Big Five personality traits.

This discrepancy hints at a peculiarity of studies that involve ratings of important others at multiple points in time: Siblings—close in age to the targets and encountering

similar developmental tasks (Luan et al., 2016)—might agree regarding the direction of personality change, whereas normative changes in the child-parent relationship might affect how parents gauge their children, potentially leading to a larger amount of disagreement between self- and parent-ratings with increasing age.

Taken together, the overall mixed findings indicate that self-reported age trends cannot be simply mapped onto the observer's perspective. Cross-sectional studies point in part toward age trends that are comparable to those observed in self-reports, but effects often appear attenuated—thus, one might expect to find maturation effects in other-reports given a sufficient sample size. However, these results should be considered preliminary because studies either sampled a very restricted age range or had comparatively small sample sizes for the age range of interest. Longitudinal studies paint an even more complex picture, which can possibly be explained by changes in the relationship between targets and raters.

Self-Other Agreement as a Domain of Maturation

Other-reports do more than merely allow researchers to test whether self-reported age effects also emerge in another point of view: When other-reports are included in addition to self-reports, agreement between self- and other-reports can be investigated as a third measure. It is important to mention that self-other agreement—which depends on the relative standing of the target among other targets in both self- and other-reports—reflects a metric that is independent of the development of the mean values of self- and other-reports, respectively, and thus has the potential to contribute additional and unique information.

From a theoretical perspective, age effects on self-other agreement appear plausible in the light of previous research on determinants of self-other agreement in personality ratings. Before we provide an overview of factors that influence self-other agreement, we will briefly present the two different ways in which self-other agreement can be conceptualized.

Trait Agreement versus Profile Agreement

Trait agreement of self- and other-reported personality is assessed as the correlation between self- and other-reports of one trait across targets. The resulting coefficients—one per trait—reflect, for example, whether targets who describe themselves as highly extraverted (relative to the self-reports of other targets) are also described as highly extraverted by others (relative to the other-reports of other targets). This approach has also been labeled "nomothetic" (e.g., in Funder, Kolar, & Blackman, 1995), and the resulting correlations allow comparisons of self-other agreement to be made between different traits.

Profile agreement is assessed by correlating self- and other-reports for one target across traits. The resulting coefficients—one per target—reflect, for example, whether targets who describe themselves as more extraverted than agreeable are also described as more extraverted than agreeable by others. In contrast to the nomothetic bivariate trait correlations, this approach has been labeled "ideographic" (e.g., in Funder et al., 1995), and the resulting correlations allow comparisons of self-other agreement to be made between different rater-target pairs. Personality profiles and the resulting profile agreement scores confound two components: a *normative profile*, which reflects the average profile, and a *distinctive profile*, which reflects the extent to which an individual deviates from this average profile. A self-reported personality profile will often resemble the normative self-reported profile; an other-reported personality profile will often resemble the normative other-reported profile. Because the normative profiles of self- and other-reports are often highly similar, even randomly chosen self- and other-reports of different targets will often yield substantial levels of profile agreement. As a consequence, more sophisticated frameworks have been developed to disentangle the two components and therefore allow for the calculation of a distinctive profile agreement that is no longer affected by normativeness (e.g., Biesanz, West, & Millevoi, 2007; Furr, 2008).

Even though trait agreement and profile agreement seem conceptually distinct, they overlap with each other. Trait agreement can be partialled into contributions of pairs of ratings; profile agreement can be partialled into contributions of single traits. The average trait agreement across traits and the average profile agreement across targets reflect the same information about self-other agreement (Allik, Borkenau, Hřebíčková, Kuppens, & Realo, 2015).

What Influences Self-Other Agreement?

Accurately judging another person's personality is not trivial: According to the realistic accuracy model (RAM; Funder, 2012; Funder et al., 1995), multiple cognitive and interpersonal processes are necessary to solve this task. The person being judged must behave in a way that is relevant to the trait of interest. This behavior must be available for the judge to observe, and the judge must detect the relevant behavior and subsequently utilize it correctly to arrive at a valid personality judgment. It is important to mention that the accuracy of self-perceptions of traits is determined by similar processes (Connelly & Ones, 2010). Thus, both self- and other-ratings of a specific trait rely on—relevant and accessible—information about the target, and both the self- and the other-rater interpret (i.e., detect and utilize) this information to arrive at a judgment.

However, different information is available to the two judges, and the same pieces of information might be interpreted and weighted differently (Vazire, 2010). Conceptually, four factors can influence the extent to which the self and the other then agree: (a) the trait to be rated because it influences which and how much relevant information is available but also because it can affect the interpretation of the information that is available to both target and rater; (b) the target because s/he affects what information is available to the rater but also because s/he accesses and interprets information in order to arrive at a self-report; (c) the rater because s/he accesses and interprets the available information; and (d) the relationship

between the target and rater because it influences what information is shared between the target and rater.[1]

Characteristics of the trait. Studies have repeatedly found that self-other agreement systematically varies between the Big Five personality traits. In their meta-analytic review of self-other agreement, Connolly, Kavanagh, and Viswesvaran (2007) reported reliability-corrected correlations ranging from .46 (agreeableness) to .62 (extraversion). Such differences have been attributed to differences in the visibility and "evaluativeness" of traits (Connelly & Ones, 2010; Funder & Colvin, 1988; John & Robins, 1993; Watson, Hubbard, & Wiese, 2000). For example, behaviors that are associated with extraversion (being talkative, asserting oneself) are directly observable, whereas signs of neuroticism (feeling anxious, becoming stressed easily) first and foremost occur within the individual; thus, self-other agreement tends to be higher for extraversion than for neuroticism/emotional stability because the quantities of relevant information available to the rater differ. Agreeableness is a highly evaluative trait; it is associated with characteristics (warmth, consideration) that are socially valued. An ego-protection motive might disrupt people's ability to accurately interpret information that indicates their standing on such traits (Vazire, 2010), lowering self-other agreement relative to more neutrally evaluated traits such as extraversion. Another reason for lower self-other agreement for evaluative traits might be that individuals have an interest in hiding behaviors that indicate low levels of an evaluative trait and highlighting behaviors that indicate high levels of an evaluative trait, leading to less genuine information about the trait level available to observers (Connelly & Ones, 2010; but see also Allik, Realo,

[1] Notice that these four factors line up with the four moderators of accurate personality judgment identified by Funder (2012): "good trait," "good target," "good rater," and "good information." However, in Funder's account, these factors are considered with respect to their effect on accuracy (i.e., the agreement between an other-rating and the real attributes of the target) and not with respect to their effect on self-other agreement (i.e., the agreement between an other-report and a self-report). Please also note that self-other agreement is neither necessary nor sufficient for accuracy: An other-report might be accurate but yield low self-other agreement because the self-report of the target is not accurate; an other-report and self-report might agree to a large extent but both be inaccurate with regard to the real personality of the target.

Mõttus, Borkenau, et al., 2010, for arguments that these differences might be attributable to differences in variance).

Characteristics of the target. Why are some people like an open book, whereas others remain sphinxlike? Human and Biesanz (2013) reviewed the literature on judgeability and identified three broad categories. Overall, good targets are psychologically well-adjusted, have a high social status, and differ from bad targets in their socialization. More precisely, good targets have high levels of self-knowledge and behave authentically; their high status results in fewer constraints, and they tend to come from family environments that tolerate and encourage emotional expression. A critical point of these characteristics is that they correspond to the desirable normative personality, which opens up a different interpretation: High levels of self-other agreement might occur when well-adjusted targets are rated because well-adjusted targets are more similar to the normative personality profile, thus being closer to what raters think others are like *in general*. However, Human and Biesanz (2011a) found that psychological adjustment increased self-other agreement even after normativeness was accounted for. They also found that, even though well-adjusted individuals also had more self-knowledge, self-other agreement was mostly enhanced through greater judgeability: Well-adjusted individuals provided others with more information about traits that were less observable.

Characteristics of the rater. As the counterpart to the "good target," the "good rater" seems to possess similar desirable features. Vogt and Colvin (2003) found that raters with a high interpersonal orientation achieved greater accuracy when judging personality on the basis of the videotaped interactions of a target. Using the same data, Colvin and Bundick (2001) reported that good raters were warm, outgoing, empathic, autonomous, egalitarian, and less hostile. Similarly, Letzring (2008) found that good judges were socially skilled, agreeable, and well-adjusted.

Again, it is important to ask whether these effects can be explained by normativeness. Well-adjusted raters might have a good understanding of what others tend to be like in general without necessarily being better at judging what makes a target distinctive. Indeed, Human and Biesanz (2011b) found that well-adjusted raters had a better understanding of what others are like in general but were not better at perceiving the ways in which targets were distinct from the normative personality.

The relationship between target and rater. Previous research has bolstered the intuitive assumption that self-other agreement depends on not just the characteristics of the rater and the target but also on characteristics of their relationship. For example, self-other agreement between previously unacquainted college roommates was found to increase over the first months of living together (Kurtz & Sherker, 2003), and students randomly assigned to discussion groups achieved higher levels of self-other agreement after meeting once a week for 7 weeks (Paulhus & Bruce, 1992). Judgments of close acquaintances were found to lead to higher levels of self-other agreement than judgments of strangers (Funder & Colvin, 1988). Married couples, in turn, were found to have higher levels of self-other agreement than dating couples or friends (Watson et al., 2000). In a meta-analysis, Connelly and Ones (2010) concluded that different tiers of intimacy resulted in different magnitudes of self-other agreement: Spouses and dating partners achieved the highest levels of self-other agreement, followed by friends, roommates, and parents, and finally by work colleagues and incidental acquaintances who only slightly outperformed strangers. A more recent study by Allik et al. (2016) confirmed the concept of qualitative tiers: Couples showed higher levels of agreement than family members, who, in turn, showed higher levels of agreement than acquaintances; whereas sheer length of acquaintance did not seem to matter in this study.

The finding that features of the relationship between target and observer influence self-other agreement can presumably be explained by differences in how much information is

shared between the target and rater (information quantity) and how useful this information is with respect to the judgment of specific traits (information quality). In support of this notion, Letzring, Wells, and Funder (2006) found that both information quantity and information quality affected self-other agreement in an experimental study in which these two aspects were systematically manipulated.

Age Effects on Self-Other Agreement

To our knowledge, only a few studies thus far have reported self-other agreement at various ages. In a cross-sectional study of 5,543 targets, Allik, de Vries, and Realo (2016) investigated various moderators of self-other agreement. Their Estonian sample covered a wide age range from 18 to 91 years (M = 46.4 years), whereas their Dutch sample consisted of psychology students with an average age of 20.2 years; targets in both samples nominated somebody who knew them well as informants. The results indicated that self-other agreement, quantified as distinctive profile correlations, was not significantly correlated with target age in either the Estonian (r = -.003) or the Dutch sample (r = .038). This study suggests that age has no prominent effect on self-other agreement. However, because these data were not collected to investigate maturation, the larger Estonian sample (N = 3,345) was spread across the entire age range and thus might not have been optimal for detecting effects in the age range in which maturation is most pronounced. The Dutch sample (N = 2,198) may have been more suitable for detecting such effects, but it covered only a rather narrow age range (mostly 18 to 24).

Luan et al. (2016) investigated self-other agreement in two longitudinal samples. In the first sample, self-parent agreement regarding emotional stability dropped below the significance threshold at age 12 but then rose again to medium levels by age 29, whereas agreement regarding agreeableness dropped at age 29. In the second sample, self-other agreement showed a broad range of patterns across age depending on the trait as well as on

the family member who delivered the rating. It is important to note, however, that agreement between self-reports and parent- or sibling-reports of personality suffers from the same confounding described earlier: Changes in self-other agreement could also be ascribed to normative changes in relationships within families.

Thus, based on the rather sparse empirical evidence, one could arrive at the preliminary conclusion that age has no systematic influence on self-other agreement. But this finding does not align with what we might predict on the basis of the literature on personality maturation and self-other agreement summarized previously. Individuals develop more mature—well-adjusted—personalities as they come of age. We expect that self-other agreement changes in synchrony with personality and thus increases from late adolescence to early adulthood, given that well-adjusted individuals in turn are easier to judge, probably have friends who are more mature themselves and are better raters of personality in general, and probably have more intimate relationships in which more relevant information is exchanged.

The Present Study

In this study, we investigated personality maturation from the perspectives of both the self and observers in a cross-sectional analysis of individuals aged 14 to 29. To our knowledge, this is the first study to compare age differences in self- and other-reported personality in a large-scale sample of more than 10,000 targets (i.e., more than 20,000 reports of personality). Furthermore, this is the first study to investigate age trends in self-other agreement covering a broad age range in a large sample of more than 200 individuals per year of age.

On the basis of theoretical considerations and the results of previous studies, we expected the following:

1. Self-reports were expected to reflect the processes of personality maturation: Older individuals were expected on average to report higher levels of emotional stability, agreeableness, and conscientiousness. Regarding the age range covered, we also expected an increase in openness to experience.

2. As suggested by previous cross-sectional studies and because self- and other-reports agree quite substantially, other-reports were expected to follow the trends shown by self-reports. However, the extent of the similarities and discrepancies between the trajectories is an open question that we attempted to address empirically.

3. Self-other agreement was expected to be positively associated with age. We had three reasons to expect this on the basis of previous research: Older targets should be "better targets" because of their greater maturity (Human & Biesanz, 2013). Because older targets are likely to have older friends, older targets should also have more mature and therefore more accurate raters (Letzring, 2008). Finally, older targets are likely to have known their raters for a longer time and might have more intimate relationships, which should again increase agreement.

Method

Participants

The data used in this study were provided by the myPersonality Project (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015)[2]. This project collected data with the help of the myPersonality app, an application on the social network site Facebook that allowed users to take psychometric tests. From 2007 to 2012, more than 7 million Facebook users completed at least one of several questionnaires that were offered. The data were gathered with an explicit opt-in consent for reuse for research purposes beyond the original project, were

[2] A number of studies using these data have been published (see mypersonality.org/wiki/doku.php?id=publications for a list). However, the topic of our manuscript does not overlap with any of these publications as none addressed age effects on personality.

anonymized, and made available for registered collaborators on the myPersonality Project homepage (http://www.mypersonality.org). Since we only reanalyzed this secondary and de-identified data, ethical approval for the present study was not required.

Users chose the length of the Big Five personality questionnaire in advance, ranging from 20 to 100 items. However, they were able to answer additional questions in blocks of 10 items each until they reached 100 items. In addition, users could send notifications to anyone on their Facebook friends list to invite them to provide other-ratings of personality. The option to provide ratings was also open to anyone on the users' friends lists. These other-reports consisted of 10 items (two items per trait) that were chosen randomly from the 100 items on the self-report questionnaires. To ensure that the other-report items had always been included on the self-reports as well, we included users' self-reports only if they had answered all 100 items. Users were given feedback on both their self-report and the average of the other-reports they had received, expressed as percentiles relative to the first 350,000 users who had completed the 100-item version of the questionnaire.

We decided to limit the sample to targets between the ages of 14 and 29 because the most prominent effects of personality maturation are supposed to occur in this period and because this decision led to an adequate sample size of $N > 200$ for each year of age. Sample correlations are subject to substantial fluctuations, but with this sample size and assuming $\rho \approx .30$-$.40$, we were able to reach a level of confidence of 80% for achieving stable estimates of self-other agreement with only small fluctuations (Schönbrodt & Perugini, 2013). The final sample consisted of 10,437 targets (5,535 women) with an overall mean age of 19.47 years (*SD* = 3.80, *Mdn* = 19, skewness = 0.74). The women were on average 19.41 years old (*SD* = 3.93, *Mdn* = 18, skewness = 0.72), and the men were on average 19.52 years old (*SD* = 3.64, *Mdn* = 19, skewness = 0.78), indicating no statistically significant age difference between the sexes, $t(10435) = 1.51$, $p = .131$.

The language of the myPersonality app was English, but access to the app was open to all Facebook users. As a consequence, the sample included users from all time zones who used more than 40 different language versions of the social network. However, the majority of the sample used an English version of Facebook (64.05% American English, 19.74% British English) and came from the time zones that cover North America (standard time zone UTC -8 to -4, 67.75%) and Europe (UTC 0 to 2, 17.11%). Language version and time zone were only negligibly confounded with age; both proxies for local background accounted for less than 2% of the variance in age. The majority of the sample (50.07%) reported being single, and 29.91% reported being in a relationship, married, or engaged. The remaining users either did not provide information on their relationship status (16.79%) or answered with one of the remaining categories such as "It's Complicated," "In an Open Relationship," or "Divorced."

Measures

Self-reported personality. Users answered a 100-item version of the International Personality Item Pool (IPIP) questionnaire that measured the Big Five personality traits (Goldberg et al., 2006). Each of the five scales consisted of 20 items (10 positively scored and 10 reverse scored) answered on a 5-point Likert scale (1 = *very inaccurate*, 2 = *moderately inaccurate*, 3 = *neither inaccurate nor accurate*, 4 = *moderately accurate*, and 5 = *very accurate*), and the Cronbach's α scale reliabilities were all satisfactory: extraversion (α = .93), emotional stability (α = .92), agreeableness (α = .88), conscientiousness (α = .92), and openness (α = .85). We allowed a maximum of two missing items per scale when calculating mean scores.

Other-reported personality. The other-report questionnaire consisted of 10 items (two per trait) chosen randomly from the 100-item version of the IPIP questionnaire employed for self-reports. Again, each item was answered on a 5-point Likert scale using the

same labels as described earlier for self-report questionnaires. Some targets were rated by multiple others, but targets were never rated more than once by the same person. For technical reasons, only the sum scores of each trait across items *and* across raters were stored, which made it impossible to assess the agreement between the two items per trait or between different raters of the same target.

On average, each target was rated by 2.45 friends (SD = 2.00, Range = 1-42). A total of 25,619 other-ratings from 25,025 unique raters were included in the data, and 531 raters contributed ratings of more than one target. We decided to ignore the dependencies among the other-ratings because (a) the majority of ratings (> 95%) were mutually independent and (b) the aggregated storing of other-reports made it impossible to recover the individual other-reports.

Linking the demographics of targets and raters. We retrieved demographic information on as many raters as possible to analyze relations between the demographics of raters and targets; 6,189 unique raters had allowed the app to access information on their age and gender. If information from multiple raters was available for one target, we randomly chose one to include in these analyses, resulting in a sample of 4,850 raters. Older targets tended to have older raters, r = .37, p < .001. However, the absolute age differences between targets and raters were non-negligible with an average of 4.56 years and showed broad variations (SD = 7.87 years). There was a weak correlation of r = .07 (p < .001) between raters' gender and targets' gender, showing a slight overrepresentation of target-rater dyads of the same gender. Including all dyads and running multilevel models resulted in virtually identical estimates of relations between targets' and raters' demographics, and therefore, simpler bivariate correlation coefficients are reported here.[3]

[3] Multilevel models are preferable in the sense that they do not require information about raters to be randomly dropped. However, there is no definite way to standardize the estimated coefficients, which renders interpretation more obscure in comparison with simple bivariate correlation coefficients.

Statistical Models

Age effects on self- and other-reports. An analysis of age effects on personality traits was computed to meet two goals: We wanted to (a) model the age effects in both self- and other-reports and (b) compare the trajectories of the self- and other-reports. Thus, we used multivariate regression analyses to simultaneously predict the self- and other-reports of each trait. The first model we ran always included a linear age effect as a predictor; the next higher order age effect (up to the cubic term) was included when the associated coefficient reached $p < .01$ for either the self- or other-report. Notice that, in these cases, the higher order terms were always included for both the self- and other-reports at the same time to allow comparisons of the coefficients between the self- and other-reports.

The comparison of the age trajectories is straightforward in the case of a simple linear model without any higher order age effects in which a simple comparison of the slopes yields the desired contrast and can easily be interpreted (e.g., Does self-reported conscientiousness increase with age as much as other-reported conscientiousness does?). However, in models with higher order terms, the slope will vary across different age values, and the comparison of regression coefficients does not have such a simple interpretation. Therefore, we compared the trajectories by jointly testing all age effects for equivalence across the self- and other-reports, thus testing the null hypothesis that self- and other-reports can both be described by the same function of age.

To test whether the cross-sectional age trajectories varied by gender, we furthermore ran additional models that included the dichotomous predictor gender (women = 0; men = 1), as well as all interactions between gender and the age terms included in the respective model.

Measures of Self-Other Agreement

Bivariate trait correlations. We calculated bivariate correlations between self- and other-reported personality traits in each of the 16 age groups separately. For subsequent

analyses, all correlation coefficients were transformed into *z*-scores with Fisher's z transformation. A measure of the mean correlation was calculated by averaging the *z*-scores of the five bivariate trait correlations. Age trends in bivariate correlations were investigated by computing linear regression models to predict the *z*-scores from age. The *z*-scores were retransformed into correlation coefficients so that they could be presented in a more common metric in the figures.

Profile correlations. We calculated profile correlations between self- and other-reported personality profiles in each of the 16 age groups separately. The overall profile correlation was computed by correlating the self- and other-reported personality scores across the five traits for each target separately. This resulted in one number per target; however, this number confounded profile similarity with normativeness. The normative profile correlation was computed by first generating normative self-report and other-report profiles (consisting of the averages of each of the five traits) for each age group and then correlating these average profiles. This resulted in only one number per age group, which reflected the normative agreement. Finally, the distinctive profile correlation disentangled profile similarity from normative agreement: The normative profiles of the respective age group were removed from a target's self- and other-report profile by traitwise standardization, separately for both types of report (e.g., Biesanz, West, & Millevoi, 2007). This resulted in distinctive profiles that reflected personality traits relative to the average of the respective age group. Correlating the distinctive self-report profile with the distinctive other-report profile for each target separately resulted in distinctive profile correlations. All correlation coefficients were transformed into *z*-scores prior to further analysis. Age trends in profile correlations were investigated by computing linear regression models that predicted the *z*-scores from age. We then retransformed the *z*-scores into correlation coefficients so that they

could be presented in a more common metric in the figures. For all analyses, only results yielding $p < .01$ were considered statistically significant to account for the large sample size.

Results

Self- and Other-Reported Trajectories

The effects of age on the self- and other-reports of the Big Five personality dimensions both for the whole sample and separately for women and men are depicted in Figure 1. As shown in Panel A of this figure, the self-reports of extraversion were generally lower than the other-reports (Cohen's $d = -0.250$,[4] see Table 1 for the corresponding descriptive statistics), but both the self- and other-reports increased linearly with age. As displayed in Table 2, the self-reports increased $b = 0.007$ scale points per year of age ($p < .001$; effect of 0.008 *SD* per year), and the other-reports increased $b = 0.010$ scale points per year of age ($p < .001$; 0.011 *SD* per year). A comparison of the slopes revealed no statistically significant difference in the linear age trend ($p = .17$). In a subsequent analysis including gender, neither the self- nor the other-reports showed significant Age × Gender interactions, suggesting similar age trajectories for women and men. However, the other-reports indicated a significant main effect of gender; that is, men scored 0.087 scale points lower (Cohen's $d = 0.10$, $p < .001$, see Tables S1-S2 and Figure 1, Panels B and C).

For emotional stability (Figure 1, Panel D), the other-reports showed no statistically significant age trend ($p = .12$). By contrast, the self-reports were best modeled as a cubic function of age (see Table 2), beginning at a lower level compared with the other-reports and increasing until they reached the level of the other-reports at about age 20. A joint comparison of the age coefficients in the cubic model indicated that the age trajectories of

[4] There are multiple ways to compute Cohen's *d* for within-subjects designs, resulting in coefficients with different interpretations. In this study, all *d*s were unpaired (as in a between-subjects design) to preserve the most common interpretation, that is, a mean difference expressed in the units of the pooled standard deviation.

self- and other-reported emotional stability differed significantly ($p < .001$). In a subsequent analysis including gender (see Table S2), neither the self- nor the other-reports showed significant Age × Gender interactions, suggesting similar age trajectories for women and men. Both self- and other-reports indicated higher levels of emotional stability for men (self-reports: $b = 0.414$ scale points, $d = -0.50$; other-reports: $b = 0.270$ scale points, $d = -0.27$, both ps $< .001$; see Tables S1-S2 and Figure 1, Panels E and F).

Self- and other-reports of agreeableness showed no linear or nonlinear age effects overall (see Figure 1, Panel G and Table 2), and a comparison of these nonsignificant coefficients indicated no differences ($p = .545$). However, the other-reports of agreeableness were generally slightly lower than the self-reports ($d = 0.077$; Table 1). The subsequent analysis including gender (Table S2) revealed a significant Age × Gender interaction ($b = -0.008$, $p = .008$): Self-reported agreeableness showed a positive age trend for women but not for men. However, no comparable gender-differentiated trend was detected for the other-reports. Furthermore, men reported lower levels of agreeableness ($b = -0.072$ scale points, $p < .001$; $d = 0.12$) and were also rated lower on agreeableness ($b = 0.125$ scale points, $p < .001$; $d = -0.15$).

Both the self- and other-reports of conscientiousness increased with age (see Figure 1, Panel J and Table 2). Whereas the self-reports showed a linear increase with no quadratic trends, the other-reports remained stable in the younger age groups at an already higher level compared with the self-reports and began to increase only around age 20, resulting in a significant quadratic term. The different age trends resulted in significant differences between the coefficients describing the self- and other-reports ($p < .001$). Subsequent analyses that included gender (Table S2) showed no significant Age × Gender interactions. Men showed lower levels of conscientiousness in other-reports ($b = -0.091$ scale points, $p < .001$; $d = 0.13$) but not in self-reports ($b = 0.010$ scale points, $p = .57$; $d = -0.02$).

Overall, self-reports of openness to experience were higher than other-reports (d = 0.47, see Table 1). Self- and other-reported openness to experience both followed a quadratic age trend (see Figure 1 and Table 2): Openness first rose quickly, but then the increase slowed down. However, we found that these trajectories were still significantly different from each other ($p < .001$): The linear increase was more pronounced in the other-reports (0.020 scale points vs. 0.012 scale points per year of age; notice that effects appear similar in magnitude when taking into account the fact that the other-reports showed a larger variance: an increase of 0.027 *SD* per year of age in other-reports vs. 0.024 *SD* in self-reports). When taking gender into account, significant Age x Gender interactions hinted that men had steeper age trajectories on both the self- and other-reports ($p = .012$, $p = .014$, see Table S1), but both interactions fell short of our significance threshold of $p < .01$.

Bivariate Trait Correlations

Across the complete sample, the self- and other-reports were moderately correlated for each of the Big Five personality traits. Extraversion showed the highest correlation (r = .41). The other traits reached lower self-other correlations: emotional stability ($r = .33$), agreeableness ($r = .34$), conscientiousness ($r = .30$), and openness ($r = .31$; all ps < .001).

Analyses of age trends in the bivariate correlations revealed no systematic changes in self-other agreement for extraversion (b = -0.001 z-scores per year, $p = .58$; see Table 3 and Figure 2). However, the correlations between self- and other-reports of emotional stability (b = 0.008 z-scores, $p = .005$) and agreeableness (b = 0.007 z-scores, $p < .001$) both increased significantly with age. Conscientiousness (b = 0.004 z-scores, $p = .206$) and openness to experience (b = 0.005 z-scores, $p = .088$) both followed the positive age trends but did not

reach the level of statistical significance. Accordingly, the mean self-other correlation aggregated across all five traits increased with age (b = 0.005 z-scores, p = .007).[5]

Quadratic and cubic age effects were not significant for any of the reported bivariate trait correlations. Furthermore, we found no convincing evidence that gender moderated the relation between self-other agreement and age (p = .54 for the mean correlation and ps > .066 for the five personality dimensions, see additional Table S2).

Profile Correlations

The average overall profile agreement across all targets was moderate to strong (r = .50, 95% CI [.493, .514]). But as expected, the normative self- and other-reported profiles were highly correlated (r = .89, bootstrapped 95% CI [.876, .910]). Thus, even randomly paired raw profiles were correlated on average. However, after accounting for normativeness, the randomly paired profiles were no longer correlated, and the distinctive profile agreement was still moderately high with an average of r = .45, 95% CI [.440, .461].

The overall correlation between self- and other-reported personality profiles increased with age (b = 0.008 z-scores per year, p < .001; see left columns of Table 4 and Figure 3, Panel A). This age effect was supplied by only the distinctive profile correlation constituent because the normative profile correlation did not change significantly with age (b = -0.015 z-scores, p = .166).[6] However, the distinctive profile correlation increased with age (b = 0.006 z-scores, p < .001): In older age groups, targets and raters showed higher agreement with respect to the ways in which the target differed from the normative personality profile.

[5] These analyses were based on only 16 data points (one per year of age), which might be considered too small for a parametric test. We thus additionally bootstrapped the coefficients and their confidence intervals by randomly drawing 10,000 samples from the original person-level data (N = 10,437) and repeating the described analyses. The bootstrapped results were similar to the parametric analyses and supported our conclusions; for example, the bootstrapped coefficient predicting the mean trait correlation from age was b = 0.0047 Fisher's z-scores per year of age, 95% CI [0.0019, 0.0074], p < .001.

[6] This analysis was again based on only 16 data points (one per year of age). As described in Footnote 5, we thus additionally bootstrapped the coefficient and its confidence interval. Results were again in line with the nonbootstrapped analysis, with b = -0.0149 Fisher's z-scores, 95% CI [-0.0333, 0.0033], p = .120.

Again, the quadratic and cubic age effects were not significant for any of the different profile correlations at the threshold we employed for the inclusion of higher order terms in our analyses ($p < .01$). However, one of the tested effects, the quadratic effect of age on the overall profile correlation, reached the significance threshold of $p < .05$ ($b_2 = -0.002$, $p = .026$). Last, we found no convincing evidence that gender moderated the relation between self-other agreement and age (ps > .058, see Table S3).

Notice that conclusions regarding age trends in the normative profile correlation, as well as the level of the overall profile correlation, depended on the coding of the Big Five personality traits. That is, if the coding of emotional stability was reversed so that higher scores reflected higher levels of neuroticism (lower levels of emotional stability), the normative profile correlations increased with age ($b = 0.042$ z-scores, $p < .001$, see right columns of Table 4 and Figure 3, Panel B): In older age groups, the average self-reported personality profile was more similar to the average other-reported personality profile. Furthermore, the overall profile correlations were generally higher, but the results for the distinctive profile correlations remained largely unaffected.

The discrepant results can be explained by a general positivity in personality profiles: When all Big Five traits were coded in the positive, socially desirable direction (i.e., the second dimension was coded as emotional stability as in Figure 3, Panel A), then a similar positivity between self- and other-reports did not contribute to the overall and normative profile correlations. However, when emotional stability was reverse-coded as neuroticism (as in Figure 3, Panel B), then a similar positivity did contribute to both the overall and normative profile correlations. Imagine a situation in which we can somehow increase the positivity of both the self- and other-reported personality profiles of one individual (i.e., we increase extraversion, emotional stability, agreeableness, conscientiousness, and openness). This would not affect the profile correlations as calculated in Panel A because increasing the

positivity of each trait simply increases the means for both the self- and other-reported Big Five factors, which are subsequently subtracted when the profile correlations are calculated. However, in the analyses underlying Panel B, increasing the positivity of the individual traits means that the levels of four of the five traits will increase, but the fifth (neuroticism) will decrease, spreading out the self- and other-reported profiles and thus increasing the profile correlation. The patterns of results reflecting increases in the normative profile correlation with age in Panel B but not in Panel A thus reflect that the normative profiles become increasingly more positive in both the self- and other-reports. However, as Panel A shows, there was no increase in the agreement regarding which "desirable" trait was more pronounced.

Additional Analyses

We ran a number of additional analyses to rule out alternative hypotheses regarding the increase in self-other agreement in both the bivariate and profile correlations. More precisely, we tested whether the trends could be explained by (a) increases in the variance of self- or other-reports, (b) increases in the reliability of self-reports, (c) increases in the number of other-report raters, or (d) increases in similarity in raters' and targets' personality.

Age-related differences in the variance of self- and other-reports. The variance of self- or other-reported measures of personality has an impact on the observed self-observer agreement because restrictions in variance lead to an underestimation of the true correlation (Allik, Realo, Mõttus, Esko, et al., 2010). In our study, psychometric features of the measures (Mõttus et al., 2015) or a selection mechanism that restricted or inflated the variance in personality measures depending on age could introduce spurious age trends in self-other correlations when they did not exist in the population. However, in line with previous findings (Mõttus, Allik, Hřebíčková, Kööts-Ausmees, & Realo, 2016), we found no evidence of systematic age trends in the variances of any of the Big Five personality measures (for

self-reports or other-reports): Regressing the scales' standard deviation on age led to coefficients that ranged from -0.001 to 0.003 scale points per year (all *p*s > .100). To test the robustness of our results, we furthermore included the standard deviations of both the self- and other-reported personality measures in the analyses in which we used age to predict the bivariate correlations. Even though the standard deviation of the self-report measures somewhat predicted the bivariate correlation coefficients between self- and other-reports for conscientiousness and openness to experience (*p*s = .014 and .028), the effects of age were not attenuated and retained their *p*-values almost perfectly.

Reliability of self-reports. Likewise, reliability sets an upper limit to the observable self-other correlations. Thus, age effects on reliability could lead to spurious age trends in self-other agreement. Indeed, we found an age trend in the reliability of self-reported emotional stability: Cronbach's α yielded values between .92 and .93 with a small increase of $b = .001$ per year ($p = .010$). No other age effects on the scale reliabilities were detected (all *p*s > .20). When we corrected the bivariate correlations between the self- and other-reports for attenuation due to measurement error (by dividing the correlation by the square root of the reliability of the self-report scale), the observed age trends did not change at all for any of the Big Five traits.

Number of raters for other-reports. As described earlier, we were not able to estimate the reliability of other-reported personality because only the summed scores had been saved. However, one obvious source of potential confounding is the number of raters. If the targets in a certain age group had systematically recruited more raters, the measurement errors in the other-reports would have decreased when the scores were averaged across multiple judges, possibly introducing spurious age trends into self-other agreement. However, the number of other-reports was not associated with age ($r = -.007$, $p = .48$).

Increasing similarity of Facebook friends. It has been suggested that self-observer agreement might be—at least in part—attributable to similarities between raters and targets. The rater might draw information from her or his own personality and assume that the target is similar (Beer & Watson, 2008). The observed increase in self-other agreement in our study could therefore be a result of increased similarity between targets and raters. Older individuals might prefer more similar friends, friends might become more similar to each other as they spend more time together, and friendships between dissimilar characters might dissolve over the course of late adolescence and young adulthood. Our data set did not contain self-reports of personality for all raters, but we were able to retrieve self-reports of raters for 1,331 targets. For 97 of those targets, we had self-reports of personality for more than one rater. In these cases, we randomly chose one of the raters to use in subsequent analyses.

The analysis of age effects on target-rater similarity followed the same strategy as the analysis of age effects on self-other agreement. Neither the bivariate trait correlations nor the mean correlations across the Big Five personality traits changed systematically with age. Coefficients ranged from -0.008 to 0.002 in *z*-scores per year (all *ps* > .10). We found no age effects on the overall profile correlations between the self-reports of targets and raters (b = -0.005 z-scores, p = .28; coding all traits in the socially desirable direction). The normative profile correlation almost *decreased* slightly with age (b = -0.006 *z*-scores, p = .013), and the distinctive profile correlation was not significantly associated with the target's age (b = -0.008 *z*-scores, p = .099). All in all, we thus found no evidence that the observed increase in self-other agreement resulted from an increase in the similarity between targets and raters.

Discussion

Age Effects on Self- and Other-Reported Personality

Older individuals describe their own personalities as more mature than younger individuals, but do observers agree with this? According to our data, they generally do, despite some differences in the timing and the magnitude of the effects.

For extraversion, positive age trends in other-reports were consistent with the respective trends in self-reports. Extraversion increased in both self- and other-reports, which might reflect the net plus in extraversion that results from substantial gains in social dominance and smaller, inconsistent changes (an early increase followed by a small decrease) of social vitality in the analyzed age range (Roberts & Mroczek, 2008). The correspondence of the age trends between self- and other-reports does not require a sophisticated explanation: Granted that both self- and other-reports are more or less noisy measures of a target's personality, age effects in both self- and other-reports should follow the same trajectories as long as there are no biases in one of the two measures that are systematically associated with age. Despite the congruence in age trends, the mean levels differed; other-reports of extraversion were consistently higher than self-reports. This could be a manifestation of a societal bias toward believing other people are more extraverted than they actually are, driven by the network extraversion bias: Feiler and Kleinbaum (2015) demonstrated that people's networks tend to be overpopulated by extraverts, which could lead to a miscalibration of the (self-)perception of extraversion.

Whereas older individuals reported increasingly higher levels of emotional stability, other-reports of emotional stability remained mostly stable. Thus, only self-reports confirmed the expected maturation. Furthermore, other-reports indicated higher levels of emotional stability than self-reports. One possible explanation is that cues that indicate neuroticism/low emotional stability such as worries and negative emotions are hard to observe (John &

Robins, 1993; Vazire, 2010). Thus, other-reports might indicate higher levels of emotional stability because observers lack cues of neuroticism; and observers might also miss positive age trends in emotional stability because they cannot directly observe how targets' negative emotions and worries become less frequent.

Surprisingly, self-reports indicated no overall increases in agreeableness. The expected increase was only visible in the female subsample. Meta-analytical findings indicate that age trends in self-reported agreeableness in the age range covered in the present study are comparably weak (Roberts et al., 2006), thus, statistical power to reliably detect such trends might have been low. Other-reports also indicated no age trends; in this sense, the two sources of information agreed with respect to age trends.

Self- and other-reported conscientiousness were both positively correlated with age as expected, but the increase in other-reports seemed to "lag behind" the increase in self-reports such that the increase began at about age 20 for other-reports rather than at age 14. A multitude of occurrences might transpire between 14 and 19 to give rise to such a pattern. For example, conscientiousness might increase early in domains that are more visible to the self than to others. It is also possible that individuals first form the intention to become more conscientious, which translates into age effects in self-reports, but reliable actions that signal conscientiousness to peers follow only after a certain amount of incubation. Apart from the age trends, individuals, especially women, consistently reported lower levels of conscientiousness than others attested to them. This might be a manifestation of the letter of recommendation effect suggested by Leising, Erbs, and Fritz (2010): Raters who like the person they are describing—as seems likely for the Facebook friends in our study—draw a more positive picture, especially on items that are evaluative.

Regarding openness to experience, age trends in self- and other-reports reflected a similar shape: Openness increased, but the trend levelled off as respondents approached the

age of 30, which reproduced the pattern reported by Roberts and Mroczek (2008). Furthermore, people overall self-reported considerably higher levels of openness than attested by their raters. Openness is often also labeled "intellect" and has been considered an instantiation of highly evaluative traits; thus, an ego-protecting bias might skew self-reports (Vazire, 2010). Notice that this might lead to the seemingly paradoxical conclusion that the evaluativeness of a trait might result in either more positive self-reports than other-reports (as for openness) or more positive other-reports than self-reports (as for conscientiousness). However, it is possible that different traits are valued distinctly when describing others' personality than when describing oneself. For example, conscientiousness might have been of little relevance for the self-concept of the targets in the context of this study, whereas openness might have been affected by an ego-protecting bias. Further research that also assesses the perceived evaluativeness of traits from different perspectives (i.e., rating oneself vs. rating others) and in different contexts of assessments (e.g., filling out questionnaires out of curiosity vs. providing ratings in a situation where the stakes are higher) could illuminate the extent to which evaluativeness can account for mean-level differences in personality ratings. It is important to note that the pattern of results was not aligned with the notion that self-reports are unequivocally more (or less) favorable than an external perspective.

Age Effects on Self-Other Agreement

Bivariate correlations. Falling mostly between .30 and .45, the bivariate correlations between self- and other-rated traits in our study are in line with results of previous studies, including meta-analyses of self-observer agreement (Connelly & Ones, 2010; Vazire & Carlson, 2010). Notice that other-reports of some targets were aggregated across multiple raters. Thus, the observed values are likely inflated compared with what would have been observed if all targets had been rated by only one rater (Back & Nestler, 2016). Yet our

values seem to fall within the lower range of self-other agreement, which can most likely be explained by the presumably low reliabilities of the two-item other-report measures.

Overall, we observed that self-other agreement increased with age and ruled out a number of alternative explanations such as age-associated changes in variance or reliability. It is interesting that self-other agreement regarding extraversion did not follow this age trend but instead began on a comparably high level and remained stable across all age groups. One possible explanation is that cues for extraversion are so easily available (Vazire, 2010) and correctly utilized that there is little space for "improvement." On a similar note, Connelly and Ones (2010) found that intimacy had a smaller influence on self-other agreement in extraversion ratings than for the other Big Five personality traits. Differences in age trends between the traits might also emerge because different chapters of a person's life emphasize the expression of different traits. For example, college as a time for socializing might offer many opportunities to learn how extraverted one's friends are, whereas later, the focus switches to traits that are more relevant to the job market or family formation.

Profile correlations. Regarding the close relation between trait agreement and profile agreement (Allik et al., 2015), one would expect distinctive profile correlations to follow the same age trends as the bivariate trait correlations. And, indeed, distinctive profile agreement increased with age: The self-reports of older respondents showed idiosyncratic patterns that better matched those reported by their observers. This effect cannot be attributed to an undifferentiated "better knowledge" of personality, but it shows that raters and targets demonstrated greater agreement with respect to what differentiates the specific target from other targets. Again, we were able to rule out alternative explanations, for example, an increased similarity between targets' and respondents' self-reports.

By contrast, the normative profile correlations did not increase when the Big Five were coded in alignment: The average self-report did not become more or less similar to the

average other-report. However, they did increase when emotional stability was recoded as neuroticism, capturing the increasing positivity of the average ratings with age due to personality maturation.

Consistency with previous findings. As mentioned earlier, Allik et al. (2016) were not able to detect a significant correlation between age and self-other agreement in a sample that was approximately half the size of ours. However, a closer look revealed that this might be due to differences in the age ranges covered and smaller sample sizes: The Estonian sample covered all ages from 18 to 91 and indeed indicated no age effects on self-other agreement at all; but the Dutch sample, consisting of psychology students (mostly between the ages of 18 and 24), showed a slight positive trend of $r = .038$ that did not reach statistical significance. This effect is actually comparable in magnitude to the respective increase of $b = 0.006$ in our study, which translates into $r = .033$.

The study by Luan et al. (2016) also did not detect consistent, statistically significant effects of age on self-other agreement; however, agreement seemed to follow different trends depending on the rater (mother, father, sibling); which might indicate that the relationship between the targets and their raters changed differentially over time. Thus, the age trends in self-other agreement in this study might be particularly informative with respect to family dynamics but less so regarding general trends in self-other agreement.

Why does self-other agreement increase with age? As predicted, we found as an overall trend that self-other agreement increased with age. We already suggested that this effect might be caused by multiple factors: targets, raters, and the relationship between targets and raters. Individuals might gain more self-insight as they age, thus delivering more realistic self-reports in the sense that they fall more in line with an external perspective. Observers might gain more insight into the target's personality, thus delivering more realistic other-reports in the sense that they fall more in line with an internal perspective. And the

relationships between older targets and their raters might yield greater quantities of higher quality information (Letzring et al., 2006). Older targets may have known their raters for a longer time. They may have spent time together in more intimate and less standardized settings than younger targets (i.e., spending evenings together vs. spending time together in school). Besides the information that is exchanged, the interpretation of the available information might become more similar with increasing age. Raters might become more emotionally invested in their relationships with the target and subsequently adopt some of the biases of the self (Vazire, 2010), but they might also gain more insight into the idiosyncratic links between a target's behavior and the underlying traits and thus interpret the information in a more valid manner.

Of course, targets, raters, and their relationships are not orthogonal entities: More mature individuals might, for example, strive for more intimate relationships that promote authentic expressions of behavior or involve more communication about one's inner world. They might also be better at integrating verbal or nonverbal feedback from others into their self-concept, leading to convergence between internal and external perspectives.

Limitations and Future Directions

Due to the cross-sectional nature of our data, we were unable to rule out the possibility that both the age trajectories we modeled and the increase in self-other agreement reflect cohort effects rather than age effects. Furthermore, self-selection into the sample could interact with age, rendering the age trends we found spurious. For example, younger Facebook users might have used the app for sheer entertainment, whereas older users (less familiar with Facebook apps) might have installed it only when they had genuine interest in getting feedback on their own personality. As a result, the older respondents might have been more invested in the questionnaires overall than their younger counterparts, and this might account for why they reached higher levels of self-other agreement. Whereas we cannot rule

out such effects, we did not find that self-reports became more reliable with increasing age, which could be one possible indicator of increasing rigor in self-reports.

Selection bias could also have arisen on the side of the rater: Younger app users might have sent invitations to distant acquaintances out of curiosity about their ratings, whereas older users might have considered the matter of personality judgments more private and may have invited only close friends to rate them. Although we again cannot rule out such effects, we found that the number of raters was not systematically associated with age, a finding that rules out certain alternative hypotheses such as younger targets inviting both close friends and distant acquaintances but older targets restricting their invitations to their dearest friends.

The pattern of results in our study—age trends in self-reports as well as the strength of the correlations between the self- and other-reports—is aligned with previous studies overall. But more data on who rated whom and why would certainly have been desirable to further investigate the effect of age on self-other agreement. Future studies could include longitudinal data to test whether self-other agreement increases within subjects. Other-reports of personality could come from the same raters on multiple occasions but also from different raters to disentangle the acquaintanceship effect from alternative underlying mechanisms. And a polydyadic study—including not only multiple raters per target but also multiple targets per rater—could yield more insight into the extent to which maturation leads to targets who are more easily rated or raters with a higher accuracy as well as insight into the unique contribution of the quality of the relationship between rater and target.

As the Big Five constitute a highly abstract level of personality description, it might be interesting to analyze whether the maturation effects found in this study can be replicated for more specific traits such as the facets of the Big Five. In the present study, other-reports of personality had unfortunately been aggregated across items and raters before being saved, thus rendering it impossible to calculate potentially more reliably item-level profile

correlations; but this limitation can easily be avoided in future studies. Furthermore, future studies could investigate agreement regarding personality traits beyond the Big Five or values, and one could investigate whether indicators of interpersonal processes such as popularity and liking become more consistent across sources of ratings with age. Extending this line of thought to meta-accuracy (Carlson, Vazire, & Furr, 2011), future studies could also examine whether and for which traits there is a systematic increase with age in the knowledge of how others see a person. Similarly, studies could be conducted to analyze whether the accuracy of personality judgments that are based on (very) brief behavioral observations—so-called thin slices of behavior—at zero acquaintance improve with age (cf. Tackett, Herzhoff, Kushner, & Rule, 2016). Furthermore, age-dependent discrepancies and agreement between explicit (i.e., self-reports) and implicit assessments of personality (e.g., Implicit Association Tests; see Back, Schmukle & Egloff, 2009; Hirschmüller, Egloff, Nestler, & Back, 2013) constitute a promising field of study for personality maturation researchers as the central question (i.e., Does the convergence of personality assessments from different sources increase with maturation?) is structurally the same (see also Nosek, 2007). Last but not least, it seems worthwhile to conduct a closer examination of the cognitive, emotional, and interpersonal (micro)processes that ultimately lead to greater agreement with age in personality ratings from the perspective of the self and others.

Conclusion

We began by asking whether raters observe the personality changes that occur from age 14 to 29 in a similar fashion as the target person and whether agreement between self- and other-reports is itself subject to age effects. To analyze these questions, we used a large data set consisting of more than 10,000 self- and other-ratings of the Big Five personality traits, providing more than 200 observations per age. In sum, our analyses showed notable commonalities between self- and other-reports: Regarding age effects in mean levels of

personality, we found that the inside perspective and the outside perspective converge for extraversion and agreeableness, and with certain qualifications also for conscientiousness and openness. Only emotional stability showed a distinct mismatch with maturation effects in self- but not in other-reports.

Bivariate trait correlations between self- and other-reports tended to increase with age with the exception of extraversion, which remained stable at a high level. Finally, overall and distinctive profile correlations both also reflected significant increases in self-other agreement with age. Thus, coming of age results in increases not only in the mean levels of certain traits but also in agreement between self-views and other-views of personality. We suggest that this finding contributes another piece that fits into the fascinating puzzle of personality maturation.

References

Allik, J., Borkenau, P., Hřebíčková, M., Kuppens, P., & Realo, A. (2015). How are personality trait and profile agreement related? *Frontiers in Psychology*, *6*, 785. doi:10.3389/fpsyg.2015.00785

Allik, J., de Vries, R. E., & Realo, A. (2016). Why are moderators of self-other agreement difficult to establish? *Journal of Research in Personality*, *63*, 72–83. doi:10.1016/j.jrp.2016.05.013

Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality*, *18*, 445–462. doi:10.1002/per.524

Allik, J., Realo, A., Mõttus, R., Borkenau, P., Kuppens, P., & Hřebíčková, M. (2010). How people see others is different from how people see themselves: A replicable pattern across cultures. *Journal of Personality and Social Psychology*, *94*, 168–182. doi:10.1037/a0020963

Allik, J., Realo, A., Mõttus, R., Esko, T., Pullat, J., & Metspalu, A. (2010). Variance determines self-observer agreement on the Big Five personality traits. *Journal of Research in Personality*, *44*, 421–426. doi:10.1016/j.jrp.2010.04.005

Allik , J., Realo, A., Mõttus, R., Pullmann, H., Trifonova, A., McCrae, R. R., & 56 Members of the Russian Character and Personality Survey. (2009). Personality traits of Russians from the observer's perspective. *European Journal of Personality*, *23*, 567–588. doi:10.1002/per.721

Back, M. D., & Nestler, S., (2016). Accuracy of judging personality. In J. A. Hall, M., Schmid Mast, & T. V., West (Eds.). The social psychology of perceiving others accurately (pp. 98-124). Cambridge University Press.

Back, M. D., Schmukle, S. C. & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97*, 533–548. doi:10.1037/a0016229

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

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. doi:10.1037/0022-3514.92.1.119

Bleidorn, W. (2015). What accounts for personality maturation in early adulthood? *Current Directions in Psychological Science*, *24*, 245–252. doi:10.1177/0963721414568662

Bleidorn, W., Klimstra, T. A., Denissen, Jaap J. A., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science*, *24*, 2530–2540. doi:10.1177/0956797613498396

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology, 101*, 831–846. doi:10.1037/a0024297

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484. doi:10.1146/annurev.psych.55.090902.141913

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In F. J. Bernieri & J. A. Hall (Eds.),

Interpersonal sensitivity: Theory and measurement (pp. 47–65). Mahwah, NJ: Erlbaum.

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. doi: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. doi:10.1037/a0021212

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, *15*, 110–117. doi:10.1111/j.1468-2389.2007.00371.x

De Bolle, M., De Fruyt, F., McCrae, R. R., Löckenhoff, C. E., Costa Jr., P. T., Aguilar-Vafaie, M. E., ... & Avdeyeva, T. V. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. *Journal of Personality and Social Psychology*, *108*, 171-185. doi:10.1037/a0038497

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69–106. doi:10.1111/j.1529-1006.2004.00018.x

Feiler, D. C., & Kleinbaum, A. M. (2015). Popularity, similarity, and the network extraversion bias. *Psychological Science*, *26*, 593–603. doi:10.1177/0956797615569580

Field, D., & Millsap, R. E. (1991). Personality in advanced old age: Continuity or change? *Journal of Gerontology*, *46*, P299–P308. doi:10.1093/geronj/46.6.P299

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

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

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, *69*, 656–672. doi:10.1037/0022-3514.69.4.656

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

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. doi:10.1016/j.jrp.2005.08.007

Göllner, R., Roberts, B. W., Damian, R. I., Lüdtke, O., Jonkmann, K., & Trautwein, U. (2016). Whose "storm and stress" is it? Parent and child reports of personality development in the transition to early adolescence. *Journal of Personality.* Advance online publication. doi:10.1111/jopy.12246

Hirschmüller, S., Egloff, B., Nestler, S., & Back. M. D. (2013). The dual lens model: A comprehensive framework for understanding self-other agreement of personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 104*, 335–353. doi: 10.1037/a0030383

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

Human, L. J., & Biesanz, J. C. (2011b). Through the looking glass clearly: Accuracy and assumed similarity in well-adjusted individuals' first impressions. *Journal of Personality and Social Psychology*, *100*, 349–364. doi:10.1037/a0021850

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. doi:10.1177/1088868313495593

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: Guilford.

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. doi:10.1111/j.1467-6494.1993.tb00781.x

Kandler, C., Kornadt, A. E., Hagemeyer, B., & Neyer, F. J. (2015). Patterns and sources of personality development in old age. *Journal of Personality and Social Psychology*, *109*, 175–191. doi:10.1037/pspp0000028

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*, 543–556. doi:10.1037/a0039210

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. doi:0.1111/1467-6494.t01-1-00005

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.doi: 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. doi:10.1016/j.jrp.2007.12.003

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. doi:10.1037/0022-3514.91.1.111

Luan, Z., Hutteman, R., Denissen, J. J., Asendorpf, J. B., & van Aken, M. A. (2016). Do you see my growth?: Two longitudinal studies on personality development from childhood to young adulthood from multiple perspectives. *Journal of Research in Personality.* Advance online publication. doi:10.1016/j.jrp.2016.03.004

McAdams, D. P., & Olson, B. D. (2010). Personality development: Continuity and change over the life course. *Annual Review of Psychology*, *61*, 517–542. doi:10.1146/annurev.psych.093008.100507

McCrae, R. R., Costa, P. T., Hřebíčková, M., Urbánek, T., Martin, T. A., Oryol, V. E.,. . . Senin, I. G. (2004). Age differences in personality traits across cultures: Self-report and observer perspectives. *European Journal of Personality*, *18*, 143–157. doi:10.1002/per.510

McCrae, R. R., Costa, P. T., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses. *Journal of Personality and Social Psychology*, *83*, 1456–1468. doi:10.1037/0022-3514.83.6.1456

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88*, 547–561. doi:10.1037/0022-3514.88.3.547

Mõttus, R., Allik, J., Hřebíčková, M., Kööts-Ausmees, L., & Realo, A. (2016). Age differences in the variance of personality characteristics. *European Journal of Personality*, *30*, 4-11. doi:10.1002/per.2036

Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-trait heterogeneity in age group differences in personality domains and facets: Implications for the development and coherence of personality traits. *PLoS ONE*, *10*, e0119667. doi:10.1371/journal.pone.0119667

Nosek, B. A. (2007). Implicit-explicit relations. *Current Directions in Psychological Science, 16,* 65–69. doi:10.1111/j.1467-8721.2007.00477.x

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660–679. doi:10.1037//0021-9010.81.6.660

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

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York: Guilford.

Pullmann, H., Raudsepp, L., & Allik, J. (2006). Stability and change in adolescents' personality: A longitudinal study. *European Journal of Personality*, *20*, 447–459. doi:10.1002/per.611

Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of Personality and Social Psychology*, *81*, 670–683. doi:10.1037/0022-3514.81.4.670

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25. doi:10.1037//0033-2909.126.1.3

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, *17*(1), 31–35. doi:10.1111/j.1467-8721.2008.00543.x.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*, 1–25. doi:10.1037/0033-2909.132.1.1

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

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*, 330–348. doi:10.1037/a0021717

Soubelet, A., & Salthouse, T. A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality*, *79*, 741–762. doi:10.1111/j.1467-6494.2011.00700.x

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, *101*, 862–882. doi:10.1037/a0024950

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*, 1041–1053. doi:10.1037/0022-3514.84.5.1041

Tackett, J. L., Herzhoff, K., Kushner, S. C., & Rule, N. (2016). Thin slices of child personality: Perceptual, situational, and behavioral contributions. *Journal of Personality and Social Psychology, 110,* 150–166. doi:10.1037/pspp0000044

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. doi:10.1037/a0017908

Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, *4*, 605–620. doi:10.1111/j.1751-9004.2010.00280.x

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, *71*, 267–295. doi:10.1111/1467-6494.7102005

Wagner, J., Ram, N., Smith, J., & Gerstorf, D. (2015). Personality trait development at the end of life: Antecedents and correlates of mean-level trajectories. *Journal of Personality and Social Psychology*, *111*, 411–429 doi:10.1037/pspp0000071

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

Table 1

Mean Scores and Standard Deviations of Self- and Other-Reports of Big Five Personality Traits and Results of Comparison Tests

	Self-reports		Other-reports		Comparison		
	M	*SD*	*M*	*SD*	*d*	*T*	*p*
Extraversion	3.31	0.79	3.52	0.85	-0.25	-23.45	< .001
Emotional stability	3.18	0.82	3.25	0.85	-0.09	-7.64	< .001
Agreeableness	3.54	0.61	3.48	0.83	0.08	6.78	< .001
Conscientiousness	3.28	0.67	3.44	0.81	-0.22	-18.49	< .001
Openness	4.02	0.51	3.73	0.72	0.47	40.22	< .001

Note. N = 10,437. d = Cohen's *d.* Possible range of the scales = 1–5.

Table 2

Results of Regression Analyses Predicting Self- and Other-Reports of the Big Five Personality Traits from Age and a Joint Comparison of the Respective Age Effects

	Self-reports		Other-reports		Comparison	
Analysis	*b*	*p*	*b*	*p*	*F*	*p*
Extraversion						
Intercept	3.3113		3.5165			
Age effect		< .001		< .001	1.91	.167
Age^1	0.0069	< .001	0.0101	< .001		
Emotional stability						
Intercept	3.2061		3.2612			
Age effect		< .001		.123	12.89	< .001
Age^1	0.0049	< .001	-0.0038	.374		
Age^2	-0.0028	.008	-0.0008	.301		
Age^3	0.0003	< .001	0.0000	.741		
Agreeableness						
Intercept	3.5365		3.4803			
Age effect		.084		.517	0.37	.545
Age^1	0.0027	.084	0.0014	.517		
Conscientiousness						
Intercept	3.2759		3.4129			
Age effect		< .001		< .001	12.97	< .001
Age^1	0.0214	< .001	0.0074	.003		
Age^2	0.0002	.604	0.0018	< .001		
Openness to experience						
Intercept	4.0374		3.7473			
Age effect		< .001		< .001	8.57	< .001
Age^1	0.0116	< .001	0.0200	< .001		
Age^2	-0.0008	.008	-0.0010	.023		

Note. N = 10,437. Higher order age effects were included as long as they yielded $p < .01$ in self- or other-reports. Age was centered at M = 19.47 years. The main effect of age refers to the joint significance test of all age coefficients included in the respective model. Comparison refers to the joint comparison of all age coefficients between self- and other-reports.

Table 3

Results of Regression Analyses Predicting Bivariate Correlations between Self- and Other-Reported Big Five Personality Traits from Age

Analysis	*b*	*p*
Extraversion		
Intercept	0.415	
Age	-0.001	.579
Emotional stability		
Intercept	0.455	
Age	0.008	.005
Agreeableness		
Intercept	0.446	
Age	0.007	< .001
Conscientiousness		
Intercept	0.353	
Age	0.004	.206
Openness to experience		
Intercept	0.390	
Age	0.005	.088
Mean correlation		
Intercept	0.412	
Age	0.005	.007

Note. Analyses used Fisher's *z*-transformed correlation coefficients for each of the $k = 16$ age groups (ranging from age 14 to 29) as dependent variable. These correlations were based on a total of $N = 10{,}437$ individuals. The mean correlation is the average of the five bivariate trait correlations. Age was centered at its mean of 19.47 years and used as a predictor variable. Higher order age effects were not significant.

Table 4

Results of Regression Analyses Predicting Profile Correlations between Self- and Other-Reported Big Five Personality Traits from Age

Profile correlation	Version 1		Version 2	
	b	*P*	*b*	*p*
Overall				
Intercept	0.554		0.670	
Age	0.008	< .001	0.012	< .001
Normative				
Intercept	1.393		1.382	
Age	-0.015	.166	0.042	< .001
Distinctive				
Intercept	0.485		0.477	
Age	0.006	< .001	0.007	< .001

Note. For Version 1, all traits were coded such that higher values reflect more desirable features. For Version 2, emotional stability was coded such that higher values reflect higher neuroticism. Total N = 10,437; for the overall profile correlations, cases were excluded when it was impossible to calculate profile correlations because either the self- or other-reports of the Big Five had a standard deviation of zero (n = 10,388 for Version 1; n = 10,397 for Version 2); analyses for the normative profile correlations were based on k = 16 age groups (ranging from age 14 to 29). Age was centered at its mean (19.47 years). All analyses used Fisher's z-transformed correlation coefficients as the dependent variable. Higher order age effects were not significant.

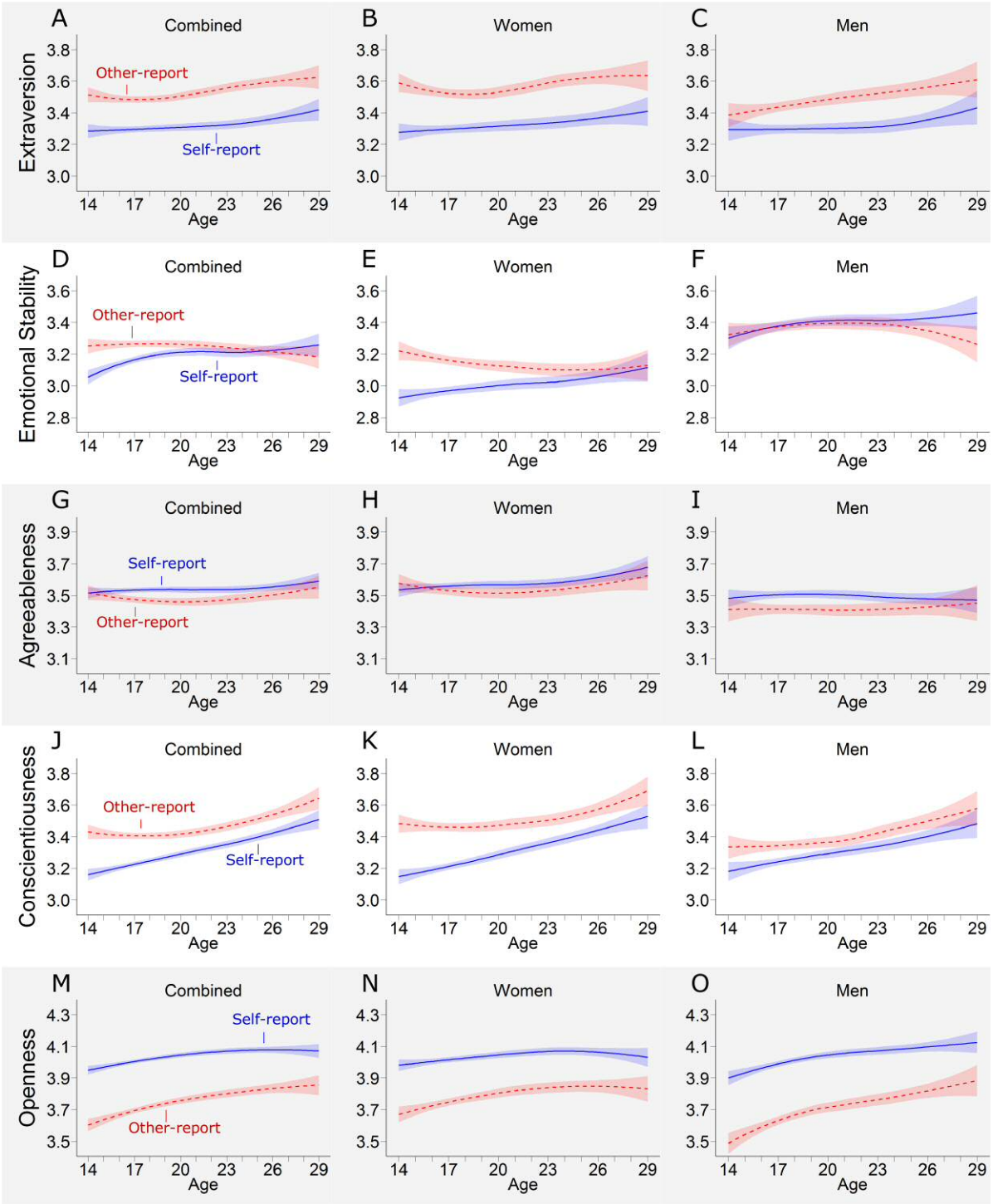

Figure 1. Cross-sectional loess-smoothed trajectories of self-reported (solid line) and other-reported (dashed line) Big Five personality traits; shaded ribbons indicate 95% confidence intervals (smoothing span $\alpha = 1$). See Table 2 for statistical models of age trends and Supplementary Table S1 for analyses including gender.

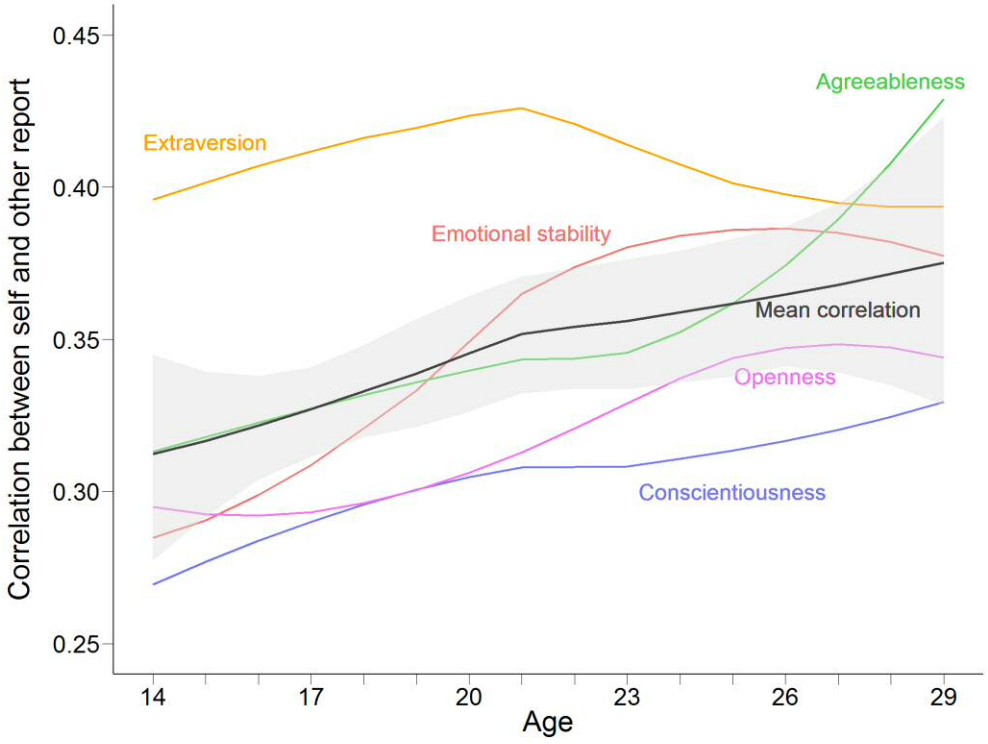

Figure 2. Cross-sectional loess-smoothed trajectories of correlations between self- and other-reported Big Five personality traits (smoothing span $\alpha = 1$). The shaded ribbon indicates the 95% bootstrapped confidence interval for the mean correlation averaged over the five personality traits (1,000 resamples). See Table 3 for statistical models of age trends.

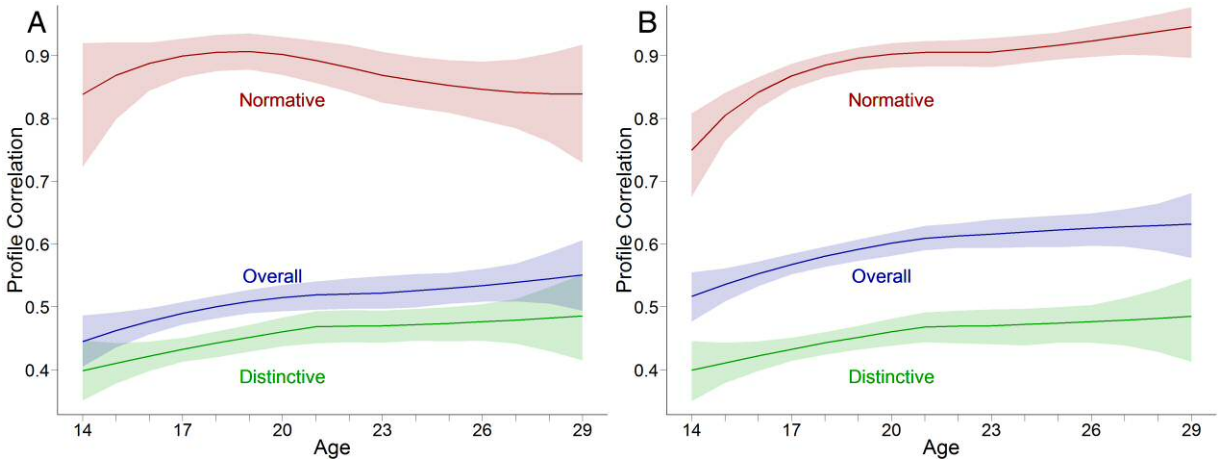

Figure 3. Cross-sectional loess-smoothed trajectories of profile correlations between self- and other-reported Big Five personality traits (smoothing span $\alpha = 1$) where all traits are coded so that higher values reflect more desirable features (Panel A) or where emotional stability is recoded so that higher values reflect higher neuroticism (Panel B). The shaded ribbons indicate bootstrapped 95% confidence intervals. See Table 4 for statistical models of age trends.

Supplementary Tables

Table S1

Mean Scores and Standard Deviations of Self- and Other-Reports of Big Five Personality Traits by Target Gender and Results of Comparison Tests

	Self-reports					Other-reports				
	Women	Men	Comparison			Women	Men	Comparison		
	M (SD)	*M (SD)*	*d*	*T*	*p*	*M (SD)*	*M (SD)*	*d*	*T*	*p*
Extraversion	3.31	3.31	0.01	0.42	.675	3.56	3.47	0.10	5.17	< .001
	(0.79)	(0.80)				(0.84)	(0.86)			
Emotional stability	2.99	3.39	-0.50	-25.64	< .001	3.14	3.37	-0.27	-13.72	< .001
	(0.78)	(0.81)				(0.84)	(0.85)			
Agreeableness	3.57	3.50	0.12	6.03	< .001	3.54	3.41	0.15	7.67	< .001
	(0.60)	(0.61)				(0.81)	(0.85)			
Conscientiousness	3.27	3.28	-0.02	-0.83	.409	3.49	3.38	0.13	6.67	< .001
	(0.67)	(0.68)				(0.80)	(0.82)			
Openness	4.03	4.02	0.02	0.93	.351	3.77	3.69	0.12	6.27	< .001
	(0.51)	(0.51)				(0.70)	(0.75)			

Note. N = 5,535 women and 4,902 men. d = Cohen's *d.* Possible range of the scales = 1–5.

Table S2

Results of Regression Analyses Predicting Self- and Other-Reports of the Big Five Personality Traits from Age and Gender

	Self-reports		Other-reports	
Analysis	*b*	*p*	*b*	*p*
Extraversion				
Intercept	3.3147		3.5572	
Gender (Male)	-0.0073	.640	-0.0869	< .001
Main effect of age		.005		.011
Age^1	0.0076	.005	0.0074	.011
Age x Gender interaction		.698		.120
Age^1 x Gender	-0.0016	.698	0.0068	.120
Emotional stability				
Intercept	2.9927		3.1241	
Gender (Male)	0.4141	< .001	0.2705	< .001
Main effect of age		.001		.013
Age^1	0.0088	.110	-0.0107	.067
Age^2	-0.0003	.782	0.0013	.207
Age^3	0.0002	.677	0.000	.888
Age x Gender interaction		.558		.013
Age^1 x Gender	-0.0071	.381	0.0157	.068
Age^2 x Gender	-0.0019	.213	-0.0031	.061
Age^3 x Gender	0.0001	.321	0.0000	.989
Agreeableness				
Intercept	3.5705		3.5388	
Gender (Male)	-0.0719	< .001	-0.1246	< .001
Main effect of age		.002		.532
Age^1	0.0064	.002	0.0018	.532
Age x Gender interaction		.008		.940
Age x Gender	-0.0083	.008	-0.0003	.940
Conscientiousness				
Intercept	3.2707		3.4597	

Gender (Male)	0.0100	.569	-0.0914	< .001
Main effect of age		< .001		< .001
Age^1	0.0243	< .001	0.0053	.110
Age^2	0.0003	.597	0.0019	.004
Age x Gender interaction		.120		.358
Age^1 x Gender	-0.0070	.099	0.0073	.152
Age^2 x Gender	-0.0001	.938	-0.0009	.400
Openness to experience				
Intercept	4.0437		3.7947	
Gender (Male)	-0.0120	.367	- 0.0931	< .001
Main effect of age		< .001		< .001
Age^1	0.0084	< .001	0.0162	< .001
Age^2	-0.0009	.036	- 0.0013	.033
Age x Gender interaction		.012		.014
Age^1 x Gender	0.0079	.013	0.0111	.014
Age^2 x Gender	0.0000	.948	-0.0001	.935

Note. N = 10,437. Higher order age effects were included as long as they yielded $p < .01$ in either the self- or other-reports in previous analyses that did not include gender. Age was centered at its mean (19.47 years). The main effect of age refers to the joint significance test of all age coefficients included in the respective model of self- or other-reports. The Age x Gender interaction refers to the joint significance test of all Age x Gender interaction terms included in the respective model of self- or other-reports.

Table S3

Results of Regression Analyses Predicting Bivariate Correlations between Self- and Other-Reported Big Five Personality Traits from Age and Gender

Analysis	*b*	*p*
Extraversion		
Intercept	.416	
Age	-.002	.550
Gender (Male)	.033	.211
Gender x Age	.003	.566
Emotional Stability		
Intercept	.319	
Age	.005	.121
Gender (Male)	.009	.714
Gender x Age	.007	.133
Agreeableness		
Intercept	.327	
Age	.011	.004
Gender (Male)	.045	.071
Gender x Age	-.009	.066
Conscientiousness		
Intercept	.311	
Age	.006	.099
Gender (Male)	-.009	.737
Gender x Age	-.007	.209
Openness to experience		
Intercept	.307	
Age	.007	.064
Gender (Male)	.023	.408
Gender x Age	-.002	.672
Mean correlation		
Intercept	.336	
Age	.006	.005
Gender (Male)	.020	.134
Gender x Age	-.002	.535

Note. Analyses used Fisher's *z*-transformed correlation coefficients for each of the $k = 32$ Gender x Age groups (ranging from age 14 to 29) as the dependent variable. These correlations were based on a total of $N = 10{,}437$ individuals. The mean correlation is the average of the five bivariate trait correlations. Age was centered at its mean of 19.47 years and used as a predictor variable.

Table S4

Results of Regression Analyses Predicting Profile Correlations between Self- and Other-Reported Big Five Personality Traits from Age and Gender

Profile correlation	*b*	*P*
Overall		
Intercept	0.573	
Age	0.007	.005
Gender (Male)	-0.041	.004
Gender x Age	0.003	.442
Normative		
Intercept	1.468	
Age	-0.006	.664
Gender (Male)	-0.026	.785
Gender x Age	-0.038	.058
Distinctive		
Intercept	0.455	
Age	0.006	.018
Gender (Male)	0.036	.009
Gender x Age	0.002	.668

Note. All traits were coded such that higher values reflect more desirable features. Total N = 10,437; for the overall profile correlations, cases were excluded when it was impossible to calculate profile correlations because either the self- or other-reports of the Big Five had a standard deviation of zero (n = 10,388 for Version 1); analyses for the normative profile correlations were based on k = 32 Age x Gender groups (ranging from age 14 to 29). Age was centered at its mean (19.47 years). The dependent variables were the Fisher's z-transformed correlation coefficients. Higher order age effects were not significant. Distinctive profile correlations were calculated by standardizing the self- and other-reported profiles *within* gender. Changes to results when standardizing across both genders affected the results only negligibly.