Age differences in the personality hierarchy: A multi-sample replication study across the life span

Yuzhan Hang, Christopher Soto, Lydia Gabriela Speyer, Liina Haring, Billy Lee, Fritz Ostendorf, René Möttus

A R T I C L E   I N F O

Keywords:
Personality development
Age differences
The Big Five domains
Facets
Nuances

A B S T R A C T

This replication and extension of Möttus and Rozgonjuk (2019) compared the extents of age-related information captured by different levels of the personality trait hierarchy (domains, facets and nuances, indexed by individual items) in several samples (N = 51,524) of different age ranges and cultural backgrounds, and tested with different instruments. Across samples and measures, lower trait hierarchy levels (especially nuances) tended to contain substantially more age-sensitive information than higher levels; most of the unique age-sensitive information was in nuances. Besides showing the need for more nuanced personality (development) research, the findings suggest ways of testing novel hypotheses that rely on systematic between-trait variance in age differences.

1. Introduction

Among other goals, personality research aims to describe and understand how personality – patterns of thinking, behaving, feeling, and motivation – change as individuals grow older. Besides rank-order stability (Roberts & DelVecchio, 2000) and associations between personality traits and individual-level variables such as life events (e.g., Bleidorn, Schwabla, & Hopwood, 2020), a commonly studied aspects of personality change is age differences mean trait levels, often described using the broad Five-Factor Model (FFM) or the Big Five trait domains: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. On average, mean scores of the personality domains gradually shift in a socially desirable direction, with people becoming more agreeable, conscientious, and emotionally stable as they age in adulthood (e.g., Caspi, Roberts, & Shiner, 2005; Donnellan & Lucas, 2008; McCrae et al., 1999; Allemand, Zimprich, & Hendriks, 2008); however, the general maturation trend can temporarily stall or even reverse in adolescence (Denissen, Van Aken, Penke, & Wood, 2013; Soto & Tackett, 2015).

Although the Big Five domains provide a useful way to summarize individual differences in behaviour, there is more to personality traits than these broad domains, as they can be split into narrower traits – personality traits form a hierarchy (Möttus et al., 2020). Facets, narrower traits below the Big Five in the trait hierarchy, contain unique personality variance above and beyond domains (Jang, McCrae, Angleitner, Riemann, & Livesley, 1998). It is therefore not surprising that facet level analyses have shown more varied developmental age trajectories compared to their domains (Lucas & Donnellan, 2009; Möttus & Rozgonjuk, 2019; Möttus et al., 2015, 2019; Soto, John, Gosling, & Potter, 2011; Terracciano, McCrae, Brant, & Costa, 2005). For example, Soto and John (2012) investigated the mean-level age trends of Big Five domains and facets in adulthood and reported longitudinal decreases in the Rumination and Depression facets of Neuroticism but increases for the Anxiety and Irritability facets. Similarly, Ashton and Lee (2016) examined age trends in HEXACO-PI-R self-reports and found that the Unconventionality facet of Openness slightly decreased over adulthood, the Inquisitiveness facet showed a modest upward trend during adulthood, and the Aesthetic Appreciation facet remained stable during middle age and then increased. Möttus and Rozgonjuk (2019) reported that facets contain over 50% more age-related variance than the Big Five domains. As summarized by Soto et al. (2011, p. 342), a “growing body of findings indicates that

* Corresponding author at: Department of Psychology, University of Edinburgh, 7 George Square, EH8 9JZ Edinburgh, UK.

E-mail address: s1886519@ed.ac.uk (Y. Hang).
conceptualizing traits at the level of Big Five facets is necessary for a full understanding of life-span age differences in personality; research at the domain level can provide a rough sketch of these differences, but not a complete picture”.

Although facets provide a more-specific picture of personality development than the Big Five domains, they are not the lowest level of the personality trait hierarchy. That is, facets themselves can be split into still-narrower traits, referred to as nuances, that represent specific personality variance not fully shared with facets (McCrae, 2015). Nuances can be indexed by single personality questionnaire items or bundles of highly similar items and are currently the lowest level of the trait hierarchy typically assessed by personality measures (Condon et al., 2020; McCrae, 2015; Mõttus, Kandler, Bleidorn, Riemann, & McCrae, 2017). It is important to realize that nuances are not merely indicators, or stylistic expressions, of domain-level and facet-level traits, but instead display distinctive properties of traits such as cross-method agreement, rank-order stability, and heritability (Mõttus et al., 2017; Mõttus, McCrae, Allik, & Realo, 2014). For example, in Mõttus et al. (2017, 2019), even when the 240 items of the NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) were residualized for the 30 facets and five domains assessed by that inventory, the remaining variance (representing unique nuances) of most items showed these properties. Nuances may also show age trends different from their domains and facets and, hence, capture unique developmental information (McCrae, 2015; Mõttus et al., 2017).

Indeed, Mõttus and Rozgonjuk (2019) “predicted” (in a statistical sense) age from the FFM domains, their 30 facets and 300 items to systematically compare the extents of age-related information captured by different levels of the personality trait hierarchy. The study used a large sample (N = 24,000), aged between 18 and 50 years with six evenly distributed and sex balanced age groups. Participants were tested with a 300-items personality questionnaire (IPIP-NEO; Goldberg, 1999), which included a shorter 120-items version of the IPIP (the 120 items were chosen from the 300 items). The 300 items collectively captured over 40% more age-sensitive information than their facets, and over 130% more information than the five broad domains, and they also contained 20% more age-relevant information than the 120 items of the shorter version of the questionnaire. Moreover, residualizing the items for their facets (and thereby also the FFM domains) had virtually no effect on how much age-sensitive information they contained, suggesting that the variability of behaviour, thinking and affect with age was mostly driven by narrow personality characteristics better captured by single items than by broader trait constructs represented by the items’ shared variance (the residuals also out-predicted the domains and facets; see Tables 1 and 2). These findings indicated that items of the same facets and domains often varied in their age differences: indeed, Mõttus and Rozgonjuk (2019) found that for half of the facets, items had significant correlations with age in both negative and positive directions.

Nuances thus contain more detailed developmental information compared to the Big Five domains and facets. In fact, it may even be that much of the age differences in the domains and facets themselves can be accounted for by the nuances that happen to be included in them. Documenting nuance-specific associations could help researchers address some currently unanswered questions (e.g. inconsistent findings across studies may result from instruments sampling different nuances) and provide them with a more complete picture of personality development, either in terms of its details (e.g., information about which specific traits develop) or general architecture (e.g., by suggesting whether the mechanisms underlying personality development are narrow and numerous or few and broadly-acting).

We realize that some readers may be sceptical about describing age differences in terms of nuances. For example, there is currently no taxonony for nuances, but merely evidence for their existence that has resulted from various item-level analyses (Mõttus et al., 2020). We think of nuances as traits beyond the few dozens of facets that have currently been outlined for the Big Five or HEXACO domains. Among others, these

### Table 1

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Age range</th>
<th>Parameter</th>
<th>Big five</th>
<th>Facets</th>
<th>Items</th>
<th>Residuals of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>M &amp; R (2019)</td>
<td>aged 18-50</td>
<td>Mean</td>
<td>0.28</td>
<td>0.44</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>EE.PIP-NEO</td>
<td>aged 16-19</td>
<td>Mean</td>
<td>0.11</td>
<td>0.11</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>CCQ (67 items)</td>
<td>aged 2-20</td>
<td>Mean</td>
<td>0.44</td>
<td>–</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>CCQ (94 items)</td>
<td>aged 2-20</td>
<td>Mean</td>
<td>–</td>
<td>–</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>BFI-2</td>
<td>aged 18-25</td>
<td>Mean</td>
<td>0.10</td>
<td>0.11</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>BFI-2</td>
<td>aged 18-50</td>
<td>Mean</td>
<td>0.13</td>
<td>0.21</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>HEXACO</td>
<td>aged 16-19</td>
<td>Mean</td>
<td>0.13</td>
<td>0.19</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td>HEXACO</td>
<td>aged 18-50</td>
<td>Mean</td>
<td>0.24</td>
<td>0.36</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>NEO-PI-R</td>
<td>aged 18-25</td>
<td>Mean</td>
<td>0.18</td>
<td>0.25</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>NEO-PI-R</td>
<td>aged 18-50</td>
<td>Mean</td>
<td>0.03</td>
<td>0.49</td>
<td>0.66</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Note. The mean and standard deviation (SD) are across 500 replications with the training sample of 75% of the total sample. The M & R dataset indicated the results from the original study.
additional traits include those that have already received considerable attention in literature and are commonly referred to in everyday language (e.g., jealousy, competitiveness, humour, loneliness, procrastination, or gratitude) but simply have not been explicitly included among the facets of the Big Five or HEXACO, even if many of them are tapped into by the individual items of these facets. Given this, there should be little revolutionary in the idea that personality trait hierarchy could be extended below the existing few dozens of facets (Condon et al., 2020): often it would mean no more than explicitly outlining many of the traits that are already measured by various inventory items and/or have already been addressed in the literature. If so, not paying attention to possible age differences in these traits would seem like a scientific neglect.

But our current analyses do not intend to comprehensively outline how nuances vary with age, although we hope the eventual need for this to be obvious. This is exactly because there is no systematic taxonomy for nuances yet, likely because researchers never saw a need for this. As a result, our current analyses mainly intend to further strengthen the case for the pursuit of developing a taxonomy for nuances and thereby also being able to properly study their development. The more robust and replicable evidence there is that facets do not capture all of the valid and potentially useful information about individual differences, including age differences, the more obvious the need to start the hard work of clearly mapping out the traits below facets becomes (Condon et al., 2020). That is, to motivate the research community to move towards a taxonomy of nuances in the first place, the need for it has to be shown with as much evidence and rigour as possible. The present paper is one key part of this effort.

We also realize that some readers may feel overwhelmed by the outlook of describing age differences in personality using dozens of traits. The level of details that individual researchers are most comfortable with is obviously a matter of personal preference (Möttus et al., 2020; Yarkoni, 2020). But it is also a matter of what research questions researchers are focusing on. For example, regardless of the number of traits used to describe how people psychologically vary with age, examining personality development at the nuance level would also allow for testing novel kinds of hypotheses that a) pertain to general principles rather than specific traits and b) therefore result in even more parsimonious findings. Specifically, if many nuance-level traits show distinctive age trends, then it will be possible to quantitatively explore systematic differences between these traits in their age trajectories and link these with other properties of the traits (Möttus & Rozgonjuk, 2019; Möttus et al., 2020). For example, researchers could (finally) formally test the hypothesis that personality development reflects psychosocial maturation (Caspi et al., 2005) by quantifying a large and diverse sample of traits in terms of both their distinctive age trends and their links to psychosocial maturity, and then testing the association between these two trait-level properties. Similarly, quantifying the extent to which different nuance traits reflect behavioural, affective, cognitive, or motivational aspects of personality would allow researchers to test how personality development is allocated across each of these four psychological domains (Wilt & Revelle, 2015). In short, examining systematic differences between traits in their age differences could be informative about general principles of personality development that cut across particular traits, broad and few or narrow and specific. This cannot be done when considering personality trait variation with age only along a few dimensions such as the Big Five.

Some readers may also think, justifiably, that increasing the number of predictors in a model is inevitably bound to increase the amount of variance that the model can explain in the sample it is fitted to, but the same result would not apply to a different sample even within the same population. This is known as over-fitting, a serious concern in statistical modelling (Seeboth & Möttus, 2018; Yarkoni & Westfall, 2017). Thus, it may seem like nuances out-predicting higher-level traits is a given. In order to address this, the present study used a combination of regularized regression and testing the models in samples other than the samples used for constructing them: this approach has been shown to give more complex model no artefactual advantage – if anything, the opposite (Seeboth & Möttus, 2018).

To our knowledge, Möttus and Rozgonjuk (2019) have provided the only previous systematic test comparing the extent of age-related information captured by different levels of the personality trait hierarchy (e.g., domains, facets, and nuances). Therefore, the present study aims to test whether the key findings of Möttus and Rozgonjuk (2019)—that personality nuances capture substantially more age-related information than facet-level traits, and that facets in turn capture more information than broad domains like the Big Five—will replicate and generalize to several different samples with a) different age ranges and b) different cultural backgrounds that c) were tested with different instruments and using different rater perspectives.

The present research thus extended that of Möttus and Rozgonjuk (2019) in four key aspects. First, it examined both youths and adults. It is conceivable that the pattern observed in adults by Möttus and Rozgonjuk does not generalize to youths; for example, it is possible that development becomes more nuanced (i.e., differentiated) only in adulthood, whereas personality may vary with age only along a few dimensions among children and adolescents. If so, there may be less of a need for developing nuanced designs to study adolescents, as opposed to adults. Second, we analysed both self-reports and parent-reports, allowing us to test whether previous findings derived from self-reports

<table>
<thead>
<tr>
<th>CCQ parameter</th>
<th>Little six/Big five</th>
<th>Facets</th>
<th>Items</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 3 to 8 years</td>
<td>Mean 0.32</td>
<td>–</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>SD 0.02</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>95% CI 0.28-0.35</td>
<td>–</td>
<td>0.51-0.56</td>
<td>0.50-0.57</td>
<td></td>
</tr>
<tr>
<td>Age 9 to 13 years</td>
<td>Mean 0.22</td>
<td>–</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>SD 0.02</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>95% CI 0.16-0.26</td>
<td>–</td>
<td>0.33-0.40</td>
<td>0.33-0.41</td>
<td></td>
</tr>
<tr>
<td>Age 14 to 20 years</td>
<td>Mean 0.08</td>
<td>–</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>SD 0.02</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>95% CI 0.04-0.13</td>
<td>–</td>
<td>0.11-0.20</td>
<td>0.12-0.20</td>
<td></td>
</tr>
<tr>
<td>Age 3 to 8 years</td>
<td>Mean –</td>
<td>–</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>SD –</td>
<td>–</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>95% CI –</td>
<td>–</td>
<td>0.55-0.60</td>
<td>0.55-0.61</td>
<td></td>
</tr>
<tr>
<td>Age 9 to 13 years</td>
<td>Mean –</td>
<td>–</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>SD –</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>95% CI –</td>
<td>–</td>
<td>0.38-0.44</td>
<td>0.38-0.44</td>
<td></td>
</tr>
<tr>
<td>Age 14 to 20 years</td>
<td>Mean –</td>
<td>–</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>SD –</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>95% CI –</td>
<td>–</td>
<td>0.17-0.24</td>
<td>0.16-0.24</td>
<td></td>
</tr>
</tbody>
</table>

Note. The mean and standard deviation (SD) are across 500 replications with the training sample of 75% of the total sample.
will generalize to informant-reports (cf. Rohrer, Egloff, Kosinski, Stillwell, & Schmukle, 2018; Kööts-Ausmees et al., 2020). It is important to study key questions of personality science using multiple methods, because any one method contains considerable degree of method-specific variance which can bias the results (Costa, McCrae, & Lückenhoff, 2019; McCrae & Möttus, 2019; Möttus et al., 2020).

Third, we tested whether our findings would generalize across different frameworks for operationalizing personality traits. This included measuring the Big Five domains with either the thirty facets of the Revised NEO Personality Inventory (NEO-FFI; Costa & McCrae, 1992; 240 items) and the Estonian Personality Item Pool (EE.PIP-NEO; Möttus, Pullmann, & Allik, 2006; 240 items) or the fifteen facets of the Big Five Inventory–2 (BFI-2; Soto & John, 2017; 60 items), as well as measuring the six HEXACO domains (Honesty/Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness; Ashton & Lee, 2020) with 25 facets (Ashton & Lee, 2007; 100 items). Beyond these hierarchically structured Big Five and HEXACO measures, we also used a less-structured measure: the common-language California Child Q-Set (CCQS; Block & Block, 1980; Caspi et al., 1992). Fourth, we tested whether our findings would generalize across different cultural contexts such as the United States, Germany and Estonia.

We expected that (a) personality nuances would contain more age-related information than facet-level traits, (b) facets, in turn, would contain more age-related information than domain-level traits, and (c) these findings would generalize across different developmental periods, rating perspectives, personality inventories, and cultural contexts. That is, we expected the original findings to replicate across the study designs and rule out the possibility that they had been driven by something questionnaire-specific, sampling or other methodological choices. We used a similar data analytic plan than the original study.

2. Method

2.1. Participants

Participants for measuring personality traits were drawn from five different studies: Ained ja Arenevad Ajud (AAA; Drugs and Developing Brains), the Common-Language California Child Q-Set Study (CCQS), the Life Outcomes of Personality Replication Project (LOOPRP), the HEXACO Online Study (HEXACOS), and the German Revised NEO Personality Inventory validation study (GNEOPIRS). Similarly to Möttus and Rozgonjuk (2019), we attempted to create a relatively uniform age and sex distribution within each sample to facilitate the comparison of results across samples. This often meant limiting the age range of the sample and dropping participants who did not fit into the age range or sampling quotas (see below).

The AAA studied drug use and mental health conditions in Estonian youths, as well as their personality traits. Participants were from Estonia and data were collected online over two years between 2018 and 2019. From among the 4,005 participants in the initial sample, (2,514 females, 1,489 males, 2 did not indicate their sex, mean age = 21.42 years), we selected participants aged between 16 and 19 years old (N = 2,269, mean age = 17.41 years; 1,304 females; 963 males, 2 did not indicate their sex, participants outside this age range were scattered across many ages; some data was dropped due to concerns with its quality). Compared to most other datasets, the questionnaire for the AAA dataset was long, consisting of a total of 240 personality questions which could have impacted participant’s willingness for carefully complete the whole questionnaire. While the GNEOPIRS questionnaire (described below) was also 240 items long, it was completed in person, thus increasing the likelihood of people completing the whole questionnaire with care.

The CCQS aimed to describe personality development from early childhood to the end of adolescence: participants were parents of 16,000 children aged from 3 to 20 years, with each parent rating one or more of their children in terms of their personality traits (Soto & John, 2014). At each age from 3 to 17, participants included exactly 500 boys and 500 girls, whereas ages 18 to 20 were combined to achieve the target of 500 boys and 500 girls. The target children were drawn from different countries with an estimated 83% from the United States, 7% from the United Kingdom or Ireland, 6% from Canada, and 4% from Australia or New Zealand. Approximately 78% of the target children were of White/Caucasian, 4% of Black/African American, 4% of Hispanic/Latino, 3% of Asian/Asian American, 1% of Native American/American Indian, 8% of mixed race/ethnicity, and 2% of another race/ethnicity. The data were collected online from a non-commercial website over ten years between 2004 and 2013 and participants volunteered to complete this survey in exchange for automatically generated feedback about their child’s personality. In the present study, target children were divided into three age groups: childhood (ages 3 to 8 years), adolescence (ages 9 to 14 years) and emerging adulthood (ages 15 to 20 years), so that non-linear age trends and age differences could be described.

The participants of the HEXACO online study (HEXACOS) completed an anonymous self-report questionnaire at the hexaco.org website between October 19, 2014 and October 18, 2018 (Lee & Ashton, 2020). There were 370,857 participants in the initial sample (mean age = 30.22 years, 57.3% were male, 41.5% were female and 1.2% did not indicate their sex). Participants were drawn from about 48 different countries (e.g. US, Germany, Australia, Japan). The HEXACOS was also divided into two age groups. In one group, we selected participants who aged between 18 and 50 years old (N = 24,000, mean age = 35.15 years) to compare with the original study. In the other group, we selected participants aged between 16 and 19 years old (N = 8,000, mean age = 17.50 years) in order to study the same age range as the AAA, for comparison. Note that there was overlap in participants’ ages and some participants may have been sampled into both the younger and older samples.

The LOOPRP estimated the replicability of the personality-outcome literature. Participants were collected from the Qualtrics Online Sample service in 2017 and quota sampling was used to ensure that the samples were representative of the United States population in terms of sex, race, and ethnicity (Soto, 2019). The 6,126 participants in the initial sample were selected into two age groups: between 18 and 50 years (N = 1,662, mean age = 35.09 years) and between 18 and 25 years old (N = 3,459, mean age = 21.89 years). There was again some overlap in participants’ ages, and therefore the samples partly overlapped.

The GNEOPIRS data came from the German Revised NEO Personality Inventory validation sample (Olaru, Schroeders, Wilhelm, & Ostendorf, 2018; Ostendorf & Angleitner, 2004). The data were collected in person over several years (1992–2002; estimated mean: 1999) in more than 50 individual studies. There were a total 12,003

---

1 Data from 364 participants with more than 40 missing personality responses were dropped (N = 3,641). Then, we limited age from 16 to 19 years, because most participants with valid data were in this age range (N = 2,269). There were 306, 919, 825, 219 participants in each age, respectively, whereas other ages had 78 people at most. We replaced missing values with the median of all values.

2 To ensure an even distribution of age and sex across the two groups, for the analysis of age differences between 18 and 50 years we sampled 4,000 participants (2,000 women, 2,000 men) based on six age levels (aged 18–25, 26–30, 31–35, 36–40, 41–45, 46–50), amounting to 24,000 participants in total; we also sampled 2,000 participants to each of four age levels (aged 16, 17, 18, 19) for the analyses of age differences between 16 and 19 years, which amounted to 8,000 participants in total. We applied a similar procedure in LOOPRP and the GNEOPIRS, trying to sample participants, so that the age ranges would be as equal as possible, although due to smaller participant numbers we could not sample equal numbers to each age level and we could not ensure sex balance.
participants (in non-clinical subsample: 35.99% were male, in clinical subsample: 50.2% were male) with ages ranging from 16 to 91 years (mean age = 30.8 years) in the initial sample. To ensure this sample was representative of the general population, it also included a clinical subsample (N = 279). Participants were from three different countries with approximately 94% from Germany, 5% from Austria and 1% from Switzerland. We separated the GNEOPIRS into two groups, with 3,324 participants aged between 18 and 50 years old in one group (mean age = 34.48) and 5,796 participants aged between 18 and 25 years old in the other group (mean age = 21.64). There was also overlap in participants’ ages in the GNEOPIRS, hence the younger and older sample partly overlapped.

We conducted a post-hoc power analysis for all included datasets using the pwr R package (Champely et al., 2018) in order to evaluate whether we had enough power to detect any true effects (Cohen, 1988), here correlations between predicted and observed ages. Results showed that sample sizes for each of our datasets had a power of 0.80 for detecting correlations of \( r = 0.10 \) or lower with two-tailed alpha set to 0.001.

### 2.2. Measures

We used five different questionnaires to measure personality traits. In the AAA dataset, participants completed a 240-item Estonian version of the International Personality Item Pool NEO (EE.PIP-NEO; Möttus et al., 2006). It mimics the structure of the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) in measuring the Big Five domains and 30 facets, with each facet assessed by eight items (McCrae & Costa, 2010). The EE.PIP-NEO was comparable to the NEO-PI-R in terms of relevant psychometric properties (Möttus et al., 2006), but the EE.PIP-NEO is linguistically simpler. Participants were required to rate the items on a 5-point Likert scale ranging from 1 (wrong/does not agree at all) to 5 (correct/completely agree).

In the CCQs, parents rated their children on the 100-item common-language California Child Q-Set (CCQ; Block & Block, 1980; Caspi et al., 1992), which includes 94 items assessing personality traits (Soto & John, 2014). Sixty seven of these 94 personality-relevant items can be aggregated to measure six broad trait domains representing the “Little Six”: the Big Five plus activity level (Soto, 2016). However, because the CCQ items were developed to be individually informative and non-redundant (Block & Block, 1980), they can also be analyzed at the item level. Participants were asked to rate each CCQ item on a 9-point Likert scale ranging from 1 (extremely uncharacteristic) to 9 (extremely characteristic).

In the LOOPRP, personality was measured using the Big Five Inventory-2 (BFI-2; Soto & John, 2017). The BFI-2 assesses the Big Five domains and 15 more-specific facets using 60 items written as short, easy-to-understand phrases, with four items per facet. Participants rated each item on a 5-point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly).

The HEXACO-PI, used to measure personality traits in HEXACOS, was developed to measure six major dimensions that have been found in several previous lexical studies of personality structure (Ashton et al., 2004; Lee & Ashton, 2004). Each HEXACO-PI domain subsamples four facets, with one additional, interstitial facet (Altruism) not scored on any of the six domains. Each facet scale includes four items, and participants rated each of the 100 HEXACO-PI items on a five-point scale from 1 (strongly disagree) to 5 (strongly agree).

In the GNEOPIRS, personality was measured with the German adaptation of the NEO-PI-R (Ollaru et al., 2018; Ostendorf & Angleitner, 2004). NEO-PI-R measures the Big Five domains and 30 facets, with each facet assessed by eight items. Participants rated the 240 German NEO-PI-R items on a 5-point scale from 0 (strongly disagree) to 4 (strongly agree).

We emphasize that none of the measures was designed to capture personality nuances, perhaps apart from CCQ. However, based on existing evidence we considered it is likely that the items of each of the measure capture a broader range of traits than they have been designed to capture — nuances — although some more than others and none of them likely as comprehensively as an instrument designed to measure nuances would do.

For simplicity, all datasets will be labelled as the name of their measurement instruments rather than the dataset names for the rest of the study. Therefore, the AAA dataset will be referred to EE.PIP-NEO, the CCQ dataset will be referred to CCQ, the LOOPRP dataset will be referred to BFI-2, the HEXACOS dataset will be referred to HEXACO, and the GNEOPIRS dataset will be referred to NEO-PI-R.

### 3. Data analysis

Statistical analyses were carried out in R (R Core Team, 2020). The EE.PIP-NEO, CCQ, BFI-2 data and the R scripts are publicly available at [the Online Supplemental Material](https://osf.io/jh4q8/); view_only=-e1ff6aa708454cb3b120db08e157a840]. Given the conditions of research ethics approval, the HEXACO data cannot be made publicly available, but it is available for researchers from Kibeom Lee and Michael Ashton. The NEO-PI-R dataset is available for researchers from Fritz Ostendorf.

To strictly test the incremental information provided by lower-level personality traits, we analysed personality nuances in terms of both raw and residualized scores. The residuals-based analyses could show how much age-related information in nuances was unique and how much was due to facets and domains they had been designed to measure. In EE.PIP-NEO, we residualized each of the 240 items for their 30 facets (and thereby the Big Five domains) using a linear regression model (with the item being residualized removed from its facet at the time). In CCQ, we residualized both the complete set of 94 personality-relevant items and the subset of 67 items assessing the Little Six domains for these domains in a similar way. In BFI-2, we residualized the 60 items for their 15 facets (and thereby the Big Five domains). In HEXACO, we also residualized the 100 items for their 25 facets and in NEO-PI-R, we residualized the 240 items for their 30 facets in a similar manner.

Following Möttus and Rozgonjuk (2019), our analyses treated age as the outcome variable (in a statistical sense) and the personality traits (raw or residualized domains, facets, and items) as the predictor variables. Specifically, age was predicted from either (a) the Big Five domains, 30 facets, 240 items, and the residuals of the EE.PIP-NEO items; (b) the Little Six domains, 94 or 67 CCQ items and their residuals; (c) the Big Five domains, 15 facets, 60 items, and residuals of the BFI-2 items; (d) the six domains, 25 facets, 100 items and residuals of the HEXACO items; or (e) the Big Five domains, 30 facets, 240 items, and residuals of the NEO-PI-R items. Since the inclusion of a high number of predictors can lead to over-fitting with complex model having an a priori advantage, (Yarkoni & Westfall, 2017; Zou & Hastie, 2005), the models were trained using (linear) elastic net regressions and applied for prediction in separate sample partitions, with a 75% — 25% split, respectively. A simple regression model trained and applied for prediction in the same sample (as is customary in most psychological modelling) would lead to inflated (over-fit) coefficients, but the current approach of using penalized regression coefficients by a) shrinking them towards zero and b) separating model training from its testing effectively mitigates this problem (Seboth & Möttus, 2016). Elastic net regressions were performed using caret (Kuhn, 2008). Models were trained using 5-fold cross-validation and fitted using an L1-L2 penalty (elastic net, lambda = 0.05; alpha = 0.5; tuning parameter = 5). In order to control for the effects of age, gender and education we used stepwise forward selection in a similar way. In BFI-2, we residualized the 60 items for their 15 facets (and thereby the Big Five domains). In HEXACO, we also residualized the 100 items for their 25 facets and in NEO-PI-R, we residualized the 240 items for their 30 facets in a similar manner.3

Note that residualizing each item for all of the facet scores does not simply reorganize the variance captured by the raw item responses. Instead, it removes some information from the items: the common variance that is shared across items. Importantly, this common variance is the main focus of classical test theory and latent trait models; thus, according to those approaches our residualizing procedure should remove all of the interesting variance from the items. However, our results show that the residualized items still retain considerable information.

---

3 Note that residualizing each item for all of the facet scores does not simply reorganize the variance captured by the raw item responses. Instead, it removes some information from the items: the common variance that is shared across items. Importantly, this common variance is the main focus of classical test theory and latent trait models; thus, according to those approaches our residualizing procedure should remove all of the interesting variance from the items. However, our results show that the residualized items still retain considerable information.
implementing the glmnet package (Friedman, Hastie, Simon, & Tibshirani, 2016), with 10-fold cross-validation and the regularization parameter (“lambda min”) minimizing the cross-validation error. This reduces the chance of capturing more variance in the outcome purely based on the inclusion of many predictors and thus helps estimating a model that is likely to also generalize to other populations (Yarkoni & Westfall, 2017). In each dataset, we repeated the training-validation procedure for 500 random sample splits. In the validation samples, the predicted ages were then correlated (using Spearman’s rho) with the actual ages. This approach yielded 2,000 correlations for each type of analysis in the EE.PIP-NEO, BFI-2, HEXACO, and NEO-PI-R — 500 for domains, 500 for facets, 500 for items, and 500 for the residuals of the items—and 1,500 correlations for each analysis in the CCQ (because there were no facets in this dataset). We then computed the mean and standard deviation for each type of correlation in each sample, and report these summary statistics as our main results. Using this procedure, if the associations between personality traits and age were primarily due to domains and facets, then the mean correlation between predicted and actual age would be stronger when age was predicted from the domains or facets than when it was predicted from the items (Seeboth & Mottus, 2018).

We note that age was predicted from these sets of predictors not because we believed age to be causally influenced by personality nor because age is inherently something worthwhile to predict from other pieces of information (although sometimes it could be, for example when age is unknown but of interest). Instead, these statistical models allowed us to estimate and compare how much age-sensitive information each set of personality traits (domains, facets, and nuances) contained. That is, prediction was used in a purely statistical sense to quantify and compare information captured across trait hierarchy levels, with no causal implications.

4. Results

4.1. Do personality nuances capture more age-related information than domain and facet traits?

In the combined age group of both the 67 items-based and 94 items-based CCQ, both age groups of the HEXACO and NEO-PI-R, the predictive accuracy of domains, facets and nuances followed the same pattern of lower-level personality traits containing more age-sensitive information than higher-order traits (see Table 1). This pattern was again displayed in the broader age group (18 to 50 years) of BFI-2. However, this pattern did not appear in the narrower age group (18 to 25 years): the accuracy of the domain-based predictions was only somewhat lower than other predictions; the accuracy of the item-based predictions and the accuracy of the facet-based predictions were almost the same. Similarly, the pattern also did not appear in the EE.PIP-NEO. The accuracies of the domain-based predictions and facet-based predictions were similar while item-based predictions were more accurate than the domain and facet-based predictions.

We also compared the predictive ability of the Big Five domains, items, and their residuals in narrower age groups within the CCQ. Specifically, we divided the CCQ into three narrower age ranges—ages 3 to 8 years, 9 to 14 years, and 15 to 20 years—and then analysed 500 random sample partitions with 75% of the age group used for training and 25% used for validation in each case. The purpose of analysing these narrower age ranges was to test whether domain, facet, and nuance-level traits could detect the non-linear trends in age differences that are often observed in childhood and adolescence (e.g., Denissen et al., 2013; Soto & Tackett, 2015). The results (Table 2) indicated curvilinear trends in the predictability of age and hence in the associations of personality traits with age: that overall amount of age-related information captured by personality traits was highest during childhood and then gradually decreased into adolescence and early adulthood, consistent with age-differences often being non-linear.

These results indicate a general trend whereby lower levels of the personality hierarchy (facets and especially nuances) contain more age-sensitive information than higher levels of the hierarchy. This pattern generalized across age groups, rating perspectives, and personality inventories, with age differences in the EE.PIP-NEO and young-adult BFI-2 sample being the exceptions, thus generally replicating and extending the findings of Mottus and Rozgonjuk (2019). However, the overall amount of age-relevant information captured by personality traits varied across both developmental periods (it was generally smaller when younger participants with smaller age ranges were studied) and instruments (with the BFI-2 containing less age-sensitive information than the HEXACO and NEO-PI-R, for example). Differences in the breadth of item content included on different measures could be one possible explanation for this finding. For example, the BFI-2 was developed to measure the most prototypical content within each Big Five domain, besides being the shortest of the measures. Its domain and facet scales may therefore be more narrowly focused and unidimensional than the other measures that we analysed, leading to less unique information at the nuance level as well as capturing less age-sensitive information at the domain level (because domains are nothing but aggregates of their nuances).

4.2. Does this pattern hold while controlling for the overlap between domain, facet, and nuance traits?

Our second set of analyses tested whether the finding that nuance-level traits capture more age-relevant information than domain- and facet-level traits would remain robust even when the overlapping information shared between domains, facets, and nuances was removed. We therefore repeated the analyses described above after residualizing each inventory item for all higher-order facets (where applicable) and domains. These residualized items no longer contained the variance of any of the Big Five (or Little Six, HEXACO) domains and facets; instead, the residuals only reflected the unique age-sensitive information captured by individual items.

In the CCQ, residualizing either 67 or 94 items did not impact their collective ability to capture age-sensitive information (see Table 1); item residuals still outperformed the domain-level traits. This also applied to the HEXACO and NEO-PI-R (see Table 1). Residualizing items for their facets had essentially no impact on how much age-sensitive information they contained, either in young adulthood specifically or across three decades of adult life more broadly. These results further suggest that age-related information was better captured by single items than by broader trait constructs measured with the items’ shared variance: age differences are mostly in nuances in which people vary even when statistically made identical (“clones”) in the domains and facets. However, in the EE.PIP-NEO and BFI-2, residualizing items for domains and facets slightly decreased their collective ability to capture age-sensitive information (see Table 1), although they still did predict age. As stated in the previous paragraph, the BFI-2 domains and facets are more narrowly focused than the other personality measured analysed in the present research, which might explain why the BFI-2 items contained less age-sensitive residual variance once the domain and facet-level information is removed. However, this hypothesis would not help explain the relatively small predictive value of residualized items in the EE.PIP-NEO.

4.3. Why items out-predicted broader traits?

The Manhattan plots (see Figs. 1–4) indicated that items of the same facets (or domains) often (but not always) differed in terms of their correlations with age, which helps to explain why item-level analyses captured more age-related information than domain and facet-level analyses. For example, while all NEO-PI-R items from the O1: Fantasy facet were negatively correlated with age, several facets such as the A6: Tender Mindedness facet and the C4: Achievement Striving facet
contained items that correlated with age in different directions, thus reducing the facet- and domain-level personality trait–age associations.

As a result, nuance-level associations can provide additional information for understanding psychological development, although we note again that items of questionnaires that have not been designed to capture nuances may not be most suitable for interpreting nuance-age correlations: they only signal the presence of such associations, but for a proper interpretation of them, appropriate instruments will ultimately be required. For example for participants aged between 18 and 50 years, NEO-PI-R items in the E1: Warmth facet that referred to liking others...
showed positive correlations with age, but items that referred to enjoying gabbing, talking, and being emotionally attached to others decreased with age: older people might be more accepting of others, but not as talkative as younger people, on average. For the E4: Activity facet, items that referred to having a fast moving life showed negative correlations with age, but items that referred to being energetic and engaging in strenuous activities showed positive correlations with age. For the C4: Achievement Striving facet, scores of items that referred to being ambitious and wanting to get ahead slightly decreased with age, whereas scores of items that referred to being enthusiastic and working excessively showed positive correlations with age: older people might not be as ambitious as younger people, on average, and yet even more hard-working.

For participants aged between 18 and 50 years, HEXACO items showed a similar pattern of occasional within-facets variance in age differences more generally and for working hard items, specifically. In the Diligence facet, the item that referred to setting ambitious goals negatively correlated with age, while scores of items that referred to working hard increased with age: older people might be more realistic in terms of setting goals than younger people, but work harder towards achieving their goals. A similar pattern for diverging age trends in ambitiousness and working hard has also been reported previously (Möttus & Rozgonjuk, 2019; Möttus et al., 2015). Such robustly divergent age trends can be very informative.

Here we aimed to explain the reasons for the generally lower predictive accuracy in domain and facet level traits in comparison to items, and therefore highlighted some item-level associations that showed opposite directions in their relations with age within the same facets. However, we want to be absolutely clear that in many cases items belonging to the same facets showed consistent correlations with age. For example, all CCQ items of the Activity domain showed negative correlations with age, while scores of items that referred to working hard increased with age: older people might be more realistic in terms of setting goals than younger people, but work harder towards achieving their goals. A similar pattern for diverging age trends in ambitiousness and working hard has also been reported previously (Möttus & Rozgonjuk, 2019; Möttus et al., 2015). Such robustly divergent age trends can be very informative.

Here we aimed to explain the reasons for the generally lower predictive accuracy in domain and facet level traits in comparison to items, and therefore highlighted some item-level associations that showed opposite directions in their relations with age within the same facets. However, we want to be absolutely clear that in many cases items belonging to the same facets showed consistent correlations with age. For example, all CCQ items of the Activity domain showed negative correlations with age, while scores of items that referred to working hard increased with age: older people might be more realistic in terms of setting goals than younger people, but work harder towards achieving their goals. A similar pattern for diverging age trends in ambitiousness and working hard has also been reported previously (Möttus & Rozgonjuk, 2019; Möttus et al., 2015). Such robustly divergent age trends can be very informative.

4.4. Do longer personality inventories provide more nuance-level information?

In their initial study comparing the age-related information captured by domain, facet, and nuance-level personality traits, Möttus and Rozgonjuk (2019) suggested that more information in terms of more variables should allow for better prediction. To test this hypothesis, we examined the age-predictive ability for personality inventories of different lengths. Specifically, we re-ordered different personality questionnaires based on the number of items in an ascending manner: BFI-2 (60 items), CCQ (67 items), CCQ (94 items), HEXACO (100 items), NEO-PI-R (240 items), and EE.PIP-NEO (240 item). In theory, the predictive ability of personality nuances should also follow this order.

In the age range between 18 and 50 years, the results obtained with the NEO-PI-R, HEXACO and the BFI-2 indeed showed that the predictive ability of the models increased with the number of items. The predictive accuracy of the 240-item NEO-P-IR was highest among the domain, facet, and nuance models, and the predictive accuracy of the 60-item BFI-2 was the lowest among these three models (see Table 1). For samples in the age range of 14 to 25 years old, the predictive ability of the NEO-PI-R, HEXACO, CCQ (67 and 94 items) and BFI-2 mostly followed the expected order (see Table 1). However, on the domain level, the predictive ability of BFI-2 slightly outperformed the predictive ability of CCQ. The only personality inventory that did not follow the general pattern of longer inventories capturing more age-relevant information was the EE.PIP-NEO. Theoretically, the predictive accuracy of the EE.PIP-NEO should have been comparable to the NEO-PI-R because they both include 240 items. However, the predictive accuracy of EE.PIP-NEO was similar to the much-shorter BFI-2.

Another noteworthy pattern apparent from these analyses is that the predictive accuracy in groups with broader age ranges generally exceeded that of groups with narrower age ranges. For example, the predictive accuracy of participants aged around 18–50 years old in the HEXACO was substantially better than the predictive accuracy of participants aged around 16 to 19 years old in the HEXACO. This same pattern can also be observed in the BFI-2 and NEO-PI-R (see Table 1). The most likely explanation for this pattern is that greater variance in age allows for greater covariance between age and personality traits.

5. Discussion

We compared the amounts of age-related information captured by domain, facet, and nuance-level personality traits. We tested whether the results generalized across samples representing different developmental periods, cultural backgrounds, rating perspectives, and personality instruments. Within each sample, we used elastic net models and random sample partitions to prevent overfitting and capitalizing on chance. Our findings support four key conclusions: (a) that lower levels of the personality hierarchy (facet and especially nuance-level traits) generally contain more age-sensitive information than higher-level trait domains like the Big Five; (b) personality nuances usually retained their age-sensitive information even after controlling for overlap with their higher-order facets and domains; (c) personality inventories that can assess a larger number of personality nuances because of containing more items generally provided more age-sensitive information than did shorter measures; and (d) the amount of age-sensitive information captured by personality inventories was generally highest during childhood, then decreased into adolescence and adulthood. Moreover, the predictive advantage afforded by nuance-level traits was substantial: across all analyses reported in Table 1 that compared the three levels of the trait hierarchy, personality nuances allowed for 36% more accurate prediction of age than did facets, and 87% more accurate prediction than did domains, whereas facets allowed for 38% more accurate predictions than did domains.

These findings successfully replicate and extend previous research on age differences in personality nuances. Using a similar analytic design in the age range of 18 to 50 years, Möttus and Rozgonjuk (2019) found that 300 items allowed for 47% and 132% more accurate predictions of age than the Big Five facets and domains, respectively. Our results similarly indicate a pervasive pattern for facets to capture more age-relevant information than domains and item-level nuances to capture more information still. Even the unique variance in items, after controlling for the
domains and facets, contained on average over 58% more age-sensitive information than the domains. This finding may seem puzzling at first: how could residualized variance, which is considered “measurement error” in classical test theory, contain much more predictive information than trait scores, which are often taken to approximate error-free “true scores”? But this can in fact be explained: what varies mostly with age in personality is the unique variance of many lower-level traits that can be loosely summarized with broader traits that aggregate these lower-level traits (as much of research does), but not accounted for by them.

This general pattern held for children, adolescents, emerging adults and adults; it also held across sample languages/cultures, rater perspectives and specific instruments. But the present results also suggest two potential boundary conditions for these findings: on average, personality nuances captured less age-sensitive information when there was less between-person variance in age (and therefore less potential for age to covary with personality nuances), and when the personality questionnaires used were relatively brief (and therefore assessed fewer personality nuances).

6. Broader implications

The present findings have important implications for understanding life-span personality differences. They suggest that age-trends in personality traits can be described along many dimensions, meaning that lower levels of the personality trait hierarchy may sometimes and possibly often be better suited for studying age differences in personality traits than higher levels. This is because these lower-level traits contain “free” information that is typically thrown away when specific personality nuances and facets are aggregated into broad, domain-level traits; in fact, the broader traits may only be useful for loosely summarizing but not for explaining age differences.

This insight implies at least three possible benefits. First, nuance-level analyses provide a possibility to describe age differences in personality traits in greater detail. Items within the same domains and facets can vary in their age trends (McCrae, 2015; Mottus et al., 2017), so it makes sense to incorporate this information to get a more accurate picture of personality age differences and avoid undue generalizations about broader traits when some of the traits’ facets and nuances show distinctive age trends. As personality research more generally matures and research questions become increasingly focused and refined, nuance-specific information may prove valuable for informing specific research questions; for example, at which life stage are people most lonely or competitive, on average. Clearly, there is still a lot to be discovered with respect to how people psychologically vary with age.

Second, representing age differences with many personality dimensions makes it possible to test new hypotheses. Specifically, nuance-level analyses will allow researchers to study systematic variations between traits in terms of how (a) nuances change with time and (b) how they intersect with developmental factors/mechanisms that characterize the nuances in different degrees. For example, some have hypothesized that personality development largely reflects social maturation (Caspi et al., 2005). To test this hypothesis, they could quantify the mean-level changes for a large pool of nuance-level traits as well as the degrees to which these traits reflect social maturity, and then correlate these two properties across all of the nuances. Similarly, some have proposed that personality maturation is largely driven by the development of self-regulation (Denissen et al., 2013). To test this hypothesis, they could correlate the degrees of age differences in a pool of personality nuances with the degrees of self-regulation required to express each nuance. Testing such associations of personality traits’ developmental trends with the traits’ other properties requires a large and diverse pool of traits, because the effective sample is the number of traits rather than the number of participants. Therefore, such tests would be best conducted when conceptualizing and measuring personality in terms of a large and diverse pool of nuance traits, rather than a much smaller set of broad traits such as the Big Five or HEXACO domains.

Third, the nuance-level analyses can help to explain why findings obtained with different instruments often do not converge (Costa et al., 2019): this is plausibly because different instruments sample different nuances. For example, age differences in Extraversion and Openness can even vary in direction, depending on which lower-level traits their measures happen to samples (Costa et al., 2019). For example, items of the diligence facet that referred to hard working positively correlated with age while scores of items that referred to having ambitious goals negatively correlated with age. This suggests that older people might have relatively more realistic goals and work hard to achieve them, while younger people might have somewhat more ambitious goals and more urge for success. Such examples of disparate correlations between items of the same facet and age were numerous.

Overall, the parsimony-focused Big Five or FFM trait models have been criticized for multiple inadequacies (e.g. lack of discriminant validity or lack of an underlying theoretical model), but they have provided the most widely used traits for operationalizing personality variance and, by virtue of allowing research findings to accumulate, they have led to fundamental advances in the description of personality-related phenomena in the past decades (Block, 2010; Mottus et al., 2020; Paunonen & Ashton, 2001). That is, the Big Five domains have been very instrumental for the progress of descriptive personality research. As a result, using nuances to describe personality variation and its links with other variables may seem to go against how many personality researchers have been trained and are used to think – to categorize lower-level traits into fewer and broader groups for greater parsimony, without necessarily considering whether those lower-level traits themselves might offer unique information. However, the present study aims to illustrate the possible value in also considering the lower-level personality traits, nuances, and not always a priori aggregating them into broader traits. We want to make it very clear that such nuanced research is not aimed at negating the Big Five-focused research, but meant to extend it and qualify its findings.

Of course, before more “nuanced” personality research can really take off, it will be necessary to develop psychometrically sound scales for measuring the nuances in future research. Developing a new generation of tools for personality assessment will require a lot of work, but it is not impossible and strategies for this have already been articulated (Condon et al., 2020; Mottus et al., 2020). However, because developing a proper model for accessing nuances is effortful and time-consuming, it is still necessary to bolster the case for this effort, as was done in the current study.

7. Limitations and future directions

Although the present findings have important implications for understanding age differences in personality traits, they should also be interpreted with four key caveats in mind. First, the idea of “predicting” age seems unusual since, implicitly, we often think of age as an explanatory variable and not an outcome. However, in the present research “prediction” was simply a tool for quantifying the amount of age-sensitive information that a set of personality traits collectively contain. Machine learning prediction models such as elastic net are particularly well suited for comparing sets of variables in how much criterion-relevant information they contain, because they allow for better control of overfitting than traditional regression models. Therefore, the present findings regarding “prediction” should only be interpreted in a statistical sense, without any implications regarding causality.

Second, one alternative explanation for the present results is that narrow personality traits within broader domains may only show distinctive age trends because they explicitly capture social expectations and roles that vary with age. However, the finding that items out-predict domains even when only closely adjacent age-levels are compared makes this unlikely to be the sole explanation: social roles and expectations are unlikely to change so quickly. The findings also generalized...
across inventories which vary in how their items tend to be written: for example, the NEO-PI-R and HEXACO items tend be more contextualized (and thereby perhaps more sensitive to age differences in social roles) than the BFI-2 items, which are more abstract and less likely to refer to any specific social roles and expectations. Moreover, BFI-2 items are based on adjectives and many items with interstitial content that fell between multiple domains or facets were excluded (Soto & John, 2017). These adjectival based items might not capture as much social role variance as contextual and behavioural items do. Moreover, the pattern also held in parent-ratings, which mitigates the possibility that the nuance-specificity of age differences may have reflected differences in how items were interpreted at different ages. Also note that personality inventories such as NEO-PI-R tend be fairly invariant across ages in how their items correlate with each other even though the items of the same traits substantially vary in how they correlate with age, suggesting that the meanings of items do not vary as much with age as the content does (e.g., Möttus et al., 2015). In our view, it is at least as plausible that personality traits per se—specific patterns of thinking, feeling, and behaving—vary in highly nuanced ways across these years.

A third caveat is that although the present study analysed data from multiple large and diverse samples of participants, all of the data were cross-sectional in nature. We therefore inferred personality change rather than directly observing it. Therefore, future research is needed to examine the development of domain, facet, and nuance-level personality traits in longitudinal samples.

Finally, although the present research operationalized personality nuances as individual questionnaire items, at present there are no theoretical models to justify which items should be used in nuance-level personality research. Most of the items analysed here—and in previous research on personality nuances—were chosen to measure broader trait domains and facets, and purposefully not to capture unique item-level information. Thus, future research would benefit from developing a stronger theoretical basis for choosing items in order to advance nuance-level analysis of age differences, as there is likely a larger pool of nuances that vary with age than is captured by the items currently used in personality inventories. Personality traits can be conceptualized at levels of abstraction other than domains, facets, and nuances, from the very broad (e.g., meta-traits) to the very narrow (e.g., highly homogenous item clusters). We have focused on the domain, facet, and item levels because they are the most prominent levels assessed by contemporary personality inventories. However, future research could establish a basis for extending the present results to other levels of abstraction. This basis could be established in ways similar to how researchers developed domain and facet-level personality measures (Möttus & Rozgonjuk, 2019). In other words, researchers should consider developing new measurement models of personality that explicitly consider nuances and therefore allow for more comprehensive personality assessment. This is not unrealistic: Condon et al. (2020) have described the concrete steps needed for the development of such models.

8. Conclusion

The present findings show that specific, nuance-level personality traits—as operationalized by individual personality questionnaire items—capture a great deal more of age-related personality information than broader traits do. We therefore conclude that future research can benefit from examining personality development at multiple levels of abstraction. Domain-level models like the Big Five and HEXACO are suitable for many purposes; for example, they allow researchers to efficiently describe broad-stroke patterns, which can be useful for describing personality development to the public or students (Möttus et al., 2020). However, nuance-level analysis can allow us to characterize age differences in personality traits in much greater detail, which will be especially useful as personality research matures and the research questions become increasingly focused. As the result of this trade-off between efficiency and specificity, some studies might like to examine personality age differences at the level of broad domains; some might want to study the extra information provided by nuances; and others might choose to address personality change using multiple levels of the personality hierarchy.

Pre-registration
The reported studies were not pre-registered.

Submission declaration
The manuscript is original, has not been previously published, and is not under concurrent consideration elsewhere.

The manuscript is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out.

The manuscript will not be published elsewhere including electronically in the same form, in English or in any other language, without the written consent of the copyright-holder, if accepted.

Funding Declaration
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Sharing Declaration
The AAA, CCQ, LOOPRP data and the R scripts are publicly available on OSF. Given the conditions of research ethics approval, the HEXACO data is not publicly available but it is available for researchers from Kiboom Lee and Michael Ashton. The GNEOPIR dataset is available for researchers from Fritz Ostendorf.

Authors’ contribution
YH contributed to conceiving the study, conducted the data analysis and drafted the manuscript.

CS contributed to collecting the CCQ and BFI-2 data which was analyzed in the current study, and provided critical feedback on drafts.

LH contributed to collecting the AAA data which was analyzed in the current study, and provided critical feedback on drafts.

FO contributed to collecting the NEO-PI-R data which was analyzed in the current study, and provided critical feedback on drafts.

LGS contributed to data analysis and provided critical feedback on drafts.

BL \ provided critical feedback on drafts.

RM contributed to conceiving the study, data analysis, and provided critical feedback on drafts.

All authors read and approved the final manuscript.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material
Supplementary data to this article can be found online at https://doi.org/10.1016/j.jrp.2021.104121.

References