Received: 13 August 2020 | Revised: 13 May 2021 | Accepted: 4 July 2021

DOI: 10.1111/jopy.12660

ORIGINAL MANUSCRIPT

WILEY

The finer details? The predictability of life outcomes from Big Five domains, facets, and nuances

Ross David Stewart[1] | **René Mõttus**[1,3] | **Anne Seeboth**[1] | **Christopher John Soto**[2] | **Wendy Johnson**[1]

[1]Department of Psychology, The University of Edinburgh, Edinburgh, UK

[2]Department of Psychology, Colby College, Waterville, ME, USA

[3]Institute of Psychology, University of Tartu, Tartu, Estonia

Correspondence
Ross David Stewart, Department of Psychology, The University of Edinburgh, Edinburgh, UK.
Email: rstewa16@ed.ac.uk

Funding information
Colby College

Abstract

Associations between personality traits and life outcomes are usually studied using the Big Five domains and, occasionally, their facets. But recent research suggests these associations may be driven by the items (reflecting personality nuances) chosen to measure these traits. Using a large dataset ($N = 6126$), we examined associations with 53 self-reported outcomes using domains, facets and items (markers for nuances), training and validating models in different sample partitions. Facets better predicted outcomes than domains (on average, 18.0% versus 16.6% of variance explained), but items provided the most accurate predictions (on average 20.9%). Removing domain and facet variance from items had no effect on their predictive validity, suggesting that outcome-related information was often in items' unique variances (i.e., nuance-specific). Item-based prediction also showed the highest discriminant validity. These observations, replicating previous findings, suggest that personality traits' valid associations with outcomes are often driven by narrow personality nuances.

KEYWORDS

facets, items, nuances, outcome, personality, traits

1 | INTRODUCTION

1.1 | Domains

Among the key topics in personality research is linking personality differences with life outcomes (Ozer & Benet-Martinez, 2006; Roberts et al., 2007). This allows researchers to assess how personality differences may impact upon people's lives, both positively and negatively, and can potentially inform interventions (Bleidorn et al., 2019). Most of this research has employed broad trait domain models such as the Five Factor Model (FFM; McCrae & John, 1992), the Big Five (Goldberg, 1990) or HEXACO (Ashton & Lee, 2020), collectively referred to as the Big Few (Mõttus et al., 2020). For example, those high in the Neuroticism domain and low in the Conscientiousness domain are more likely to make poorer health choices such as smoking and drug abuse (Bogg & Roberts, 2004), whereas those high in the Openness and Agreeableness domains are more likely to behave pro-environmentally (Soutter et al., 2020) and those high in the Extraversion domain show a higher tendency to stick to political party loyalties (Bakker et al., 2016). Such associations, often highly replicable (Soto, 2019), are so well documented that we do not need to review them here in more detail.

But despite the undeniable value of the Big Few-outcome associations, aided by their parsimony and intuitiveness, this research also has limitations. For example, the associations are generally small in magnitude, particularly when the outcomes are not broad personality-like constructs themselves such as subjective well-being or self-esteem. Although small effect sizes are recognized as a general feature of psychological research, possibly for very good reasons (Goetz et al., 2020), the overall degrees to which outcomes are linked with personality differences may sometimes be underestimated because researchers have rarely explored ways of maximizing personality traits' predictive accuracy; this may require looking beyond the Big Few alone and focusing on the combined predictive value of many traits (Mõttus et al., 2020). Although the best possible prediction from a collection of variables does not necessarily correspond to the strongest possible individual trait-outcome associations because traits partly overlap, often it is exactly the overall degrees to which personality traits and outcomes are linked that researchers care about, rather than the effect sizes of individual traits.

Moreover, the Big Few-based relations often lack specificity (discriminant validtity), as disparate outcomes show similar combinations of trait-associations (Seeboth & Mõttus, 2018). Specifically, there is a pervasive pattern of positive correlations between "success" outcomes and "positive" trait domains (Agreeableness, Conscientiousness, Extraversion, Openness and Emotional Stability; Allik et al., 2010) and negative correlations between those same domains and less desirable outcomes. For example, most of the Big Few domains are similarly associated with such diverse outcomes as longevity (Graham et al., 2017), neighborhood quality (Jokela, 2009) and pro-environmental behavior (Soutter et al., 2020). Arguably, at least in some situations personality trait-outcome associations could be more meaningful if they better distinguished among outcomes, beyond just being linked with them. Again, this may require looking beyond the Big Few alone, in particular because, given the magnitudes of the domain-outcome associations and the pervasive intercorrelations among the domains themselves, these patterns may sometimes do little more than reflect the among-domain and among-outcome correlations.

1.2 | Facets

Personality traits are organized hierarchically (Condon et al., 2020), with the Big Few domains subsuming narrower characteristics such as aspects (DeYoung et al., 2007) or facets (Ashton & Lee, 2020; McCrae & Sutin, 2018). Given that the LOOPR (Soto, 2019) data set used here was only designed to measure domains and facets, we do not focus on the aspect-traits here. Facets are not just stylistic variations in how people express their personality domains (McCrae & Costa, 2010), but traits in their own rights, with variances that are uniquely heritabile and converge across methods (Jang et al., 1998; Mõttus et al., 2014). To improve personality traits' predictive power and specificity, therefore, many researchers use facets to explore which parts of broader trait domains have stronger outcome-associations. Indeed, domains' associations may often conceal stronger and more diverse association patterns if facets of the same domain have different associations with an outcome, especially when in opposite directions (Mõttus, 2016).

For example, the Deliberation and Order facets of the Conscientiousness domain had stronger associations with Task Performance in workplace environments, whilst Dutifulness and Self-discipline facets had stronger associations with Job Dedication; facets explained up to 24% more variance in job performance than the Conscientiousness domain (Dudley et al., 2006; for similar observations see Judge et al., 2013). Likewise, subjective well-being was more strongly associated with facets than domains, with up to 15% more explained variance (Anglim & Grant, 2016). As another example, Vainik and colleagues (2019) observed that facets could explain over 400% more variance in body mass index (BMI) than domains.

Not considering facets may not only limit personality traits' predictive power and their ability to distinguish between outcomes, but also confound researchers' ability to interpretate the associations. For example, when only some facets of a domain are linked with an outcome, is it appropriate to generalize the associations to the whole domain or should researchers focus their interpretations on the specific facets involved (Mõttus, 2016)? The choice may depend on individual researchers' preferences for the most "pleasing" solution to the bandwidth-fidelity dilemma (Cronbach & Gleser, 1957; Yarkoni, 2020), but it can only help to make well-informed decisions about it when the ways in which associations generalize across facets to their domains are well documented.

1.3 | Nuances

But facets are not the most specific units of personality variance. They consist of partly overlapping, partly distinct narrow traits, nuances that are not exchangeable and may be combined in different ways to form different facets (Condon et al., 2020; McCrae, 2015).

Condon and colleagues (2020) defined nuances as "the lowest level at which patterns of responses to questionnaire items continue to have reliable specific variance" (p. 925). Often, individual items can be considered as measures of unique nuances, as most items contain reliable specific variance. However, definitionally the nuance level is broader than that of individual items as there are also likely to be sets of items that contain no reliably distinct information from one another and thus can be considered alternative indicators of the same nuance. For example, the Trust facet of Big Five Inventory (Soto & John, 2017) contains four items. The first three, "Tends to find fault with others", "Has a forgiving nature, and "Is suspicious of others' intentions" likely each measure a different nuance that we could label "judgemental", "forgiving" and "suspicious", but the fourth item, "Assumes the best about people" may well reflect (in reverse) the same nuance as the third ("suspicion").

Evidence supporting nuances as unique units of personality assessment comes from observations that individual items of a personality inventory, intended to represent specific behavioral, cognitive, motivational and affective aspects of personality, often meet formal criteria of personality traits such as agreement among raters (Mõttus et al., 2014), stability across time, and heritability (Mõttus, Kandler, et al., 2017; Mõttus et al., 2019), beside distinct developmental trends (Mõttus & Rozgonjuk, 2021) and cultural differences (Achaa-Amankwaa et al., 2020). This is true even when the variance that nuances share (i.e., the variance that constitutes facets and domains) has been removed, suggesting that many if not most items represent unique traits beyond the facets and domains they were intended to indicate.

Unsurprisingly, then, nuances are often uniquely related to outcomes (e.g., Elleman et al., 2020; Mõttus, Bates, et al., 2017; Revelle et al., 2021; Wessels et al., 2020). Sometimes only a few nuances drive the observed associations (e.g., nuances somehow similar to the outcome), such as two eating-related nuances in the Impulsivity facet driving the association with Neuroticism and BMI (Terracciano et al., 2009). In many cases, however, many more nuances are involved in the associations; it is just that how they are related to the outcomes is not well aligned with how they are typically aggregated into facets and domains. For example, income disparity was positively associated with Openness and negatively associated with Agreeableness and Conscientiousness (Elleman et al., 2020). However, it was items referring to liberal beliefs, anti-authoritarian attitudes, beliefs in self-exceptionalism and lower concern for abiding by rules that were most strongly associated with income disparity, suggesting that nuances beyond domains can contribute to understanding the outcome.

In what may be the most thorough study on the incremental predictive value of nuances yet, Seeboth and Mõttus (2018) used the National Child Development Study, with over 8000 UK participants, assessed with 50 items from the International Personality Item Pool (Goldberg, 1999). Forty outcomes were considered, such as use of the internet, racial attitudes, and income. Most associations were longitudinal, with personality traits measured five years prior to outcomes. On average, the collective associations of life outcomes with the 50 items, reflecting nuances, were around 30% stronger than associations with domains; items out-predicted domains for 37 of the 40 outcomes. The incremental predictive value of nuances was not due to model over-fitting (simulations indicated that had the associations been driven by domains, domain-based models would have had stronger associations than nuance-based models) nor content overlap (no outcome had directly corresponding nuances). However, this study did not consider facets, so it remains unclear how much of the nuances' incremental predictive value over domains could be accounted for by the unmeasured facets.

Not all studies have provided evidence for nuances' incremental associations. Mõttus, Bates and colleagues (2017) observed no significant incremental correlations of items with life satisfaction. Life-satisfaction is a very broad outcome; it is possible that personality trait-outcome associations may be strongest when the breadth of the personality measure matches that of the outcome (Asendorpf et al., 2016). However, Seeboth and Mõttus (2018) found that both broad and narrow outcomes were most strongly predicted by items to similar degrees. Alternatively, non-significant item-outcome associations, after controlling for domains, may not mean lack of relations per se: it is possible that, rather than an outcome's relations with personality traits being at the level of broad domains, they may pertain to *very many* individual nuances, so that the individual associations are too small to be statistically significant (Seeboth & Mõttus, 2018). If so, this would mirror prevalent observations in the genetics of many behavioral traits (Plomin & von Stumm, 2018): variation in each of them can be correlated with many thousands of genetic variants, so that even with extremely small and mostly statistically non-significant individual effect sizes, collectively they account for a substantial proportion of trait variance. This would also be consistent with the increasing realization that most effect sizes are indeed very small in psychological research (Goetz et al., 2020). Such small associations can be addressed with large samples and, as one solution, by testing nuances' collective associations with outcomes rather than focusing on individual trait-outcome associations.

1.4 | Disadvantages and benefits of facet- and nuance-specific analyses

Linking outcomes with facets and especially nuances means forgoing the parsimony and simplicity of the domains-based analyses, which may be prohibitive in some cases such as for public engagement, as well as off-putting for researchers with strong a priori preferences for parsimony. Also, there is currently no commonly agreed facet taxonomy for the Big Few traits, with different instruments such as BFI-2 (Soto & John, 2017), NEO Personality Inventories (McCrae & Costa, 2010), and HEXACO including different sets of facets for the same domains, or at least labeling them differently. Even more worryingly, there has been almost no work on taxonomizing and measuring nuances yet (Condon et al., 2020), hampering accumulation of a comparable body of research on their links. In fact, research into the possible value of nuances has so far only used items of existing questionnaires designed for other purposes. Observations from such work may not only be poorly comparable across studies that use different questionnaires, but also underestimate the incremental value of nuances: after all, Big Few questionnaires have been carefully designed *not* to measure anything but the core features of the Big Few domains (and their facets where applicable) and not to cover the full range of possible nuances within these domains or beyond them. So, even evidence that *some of their items contain* unique nuance information has been "against the odds".

However, that something is not yet available does not mean that it cannot be in principle, or that it should not be. As ever more of such against-the-odds evidence accumulates for nuances' potential incremental value for various purposes, the stronger the case becomes for ultimately developing nuance-level measurement models: if items *not* designed to measure specific facets and domains predict outcomes better than those facets and domains, it is likely that yet more accurate prediction could be achieved with bespoke assessments of nuances. In fact, a call for developing bottom-up taxonomies and assessment tools for nuances has already been issued (Condon et al., 2020): a widely agreed taxonomy of nuances is no less possible in principle than a widely used Big Few taxonomy, even though we may not be accustomed to think in these terms yet. Although the universe of nuances is potentially large, it is plausible that a manageable pool of them, say between 100 and 200, will provide a sufficiently useful sample (Condon et al., 2020) that could be widely used to produce just as cumulative research as we have now for broad personality domains. As a corollary, a consensual and useful taxonomy of nuances will also help to systematize the content of facets and domains, thereby helping to mitigate their jingle-jangle confounds and potentially increasing their predictive value as well (Condon et al., 2020). That is, a consideration of more nuanced associations would not replace work based on domains and facets, but supplement and even improve it.

Facet- and nuance-level observations could help research in several ways, especially once such associations are documented with bespoke and widely used measures. As discussed, they can provide more accurate assessments of the overall predictive powers of personality traits for life outcomes (Mõttus et al., 2019, 2020), not to mention better prediction of those outcomes. Individual differences researchers take great pride in this power already, because it is a unique role of this field (e.g., Deary et al., 2010; Roberts et al., 2007). However, surprisingly few studies to date have systematically explored the upper limits of this power, for example by employing modern machine learning principles; in fact, most of the existing research is not even predictive but correlational, where the "predicted" observations are used in model creation (Mõttus et al., 2020). Only deliberately attempting to maximize prediction while mitigating challenges such as over-fitting by not using the predicted observations in model creation (Yarkoni & Westfall, 2017) can show how strongly personality differences may truly matter beyond personality traits themselves. Maximizing prediction is also desirable when personality traits are used for consequential decisions such as employee selection (Lievens, 2017) or marketing decisions (Matz et al., 2016): there is no reason not to use nuances for these purposes where they demonstrably improve predictive accuracy.

Secondly, facets and nuances can help to explain how individuals scoring similarly in a domain may end up with different outcomes; that is, they can better discriminate among outcomes. Neuroticism, for example, has been associated with BMI (Sutin & Terracciano, 2016), but also with Aggressiveness (Barlett & Anderson, 2012). At the facet level, however, the two outcomes have different associations: for BMI, the only relation is often with the Impulsiveness facet of the Neuroticism domains (Sutin et al., 2011), while for Aggression the strongest relation tends to be with the Angry Hostility facet of the same domain (Jones et al., 2011). Individuals differing considerably in just these two facets could have similar domain scores, but experience different BMI and aggressiveness "outcomes." Furthermore, overlap can occur between the facets of different traits, showing how combinations of facets can have an increased predictive power. For example, Aggresiveness has also been linked with Agreeableness facets (e.g., Altruism and Compliance) (Dam et al., 2018), as well as with facets such as Warmth (Extraversion domain) and dutifulness (Conscientiousness domain) (Jones et al., 2011).

Third, knowing when smaller elements within domains drive the domain's associations with outcomes and which ones those are can help us to interpret the domain-level associations more accurately. For example, that specifically the Impulsivity facet drove Neuroticism's correlation with BMI suggests that it is not some general tendency toward negative emotionality and/or emotional instability that can contribute to being overweight but a lack of constraint, perhaps even primarily in food consumption; if so, this is hardly breaking news (but see Arumäe et al., 2020, on reverse causality).

Fourth, and relatedly, observations based on facets and nuances can shed light on the general structural architecture of personality trait-outcome associations (Mõttus et al., 2020; Seeboth & Mõttus, 2018). As one possibility, they may generally be driven by a few broad mechanisms such as those that also contribute to facets and nuances coalescing into domains: this "latent trait" scenario, probably what many studies intuitively assume, would be consistent with facets and nuances *not* predicting outcomes better than domains. If interventions are envisioned, they could then well target the broad domains (Bleidorn et al., 2019). As another possibility, however, trait outcome-associations may be largely driven by numerous specific mechanisms, as would be evidenced by facets and especially nuances out-predicting domains. Strikingly, Seeboth and Mõttus (2018) found that removing the Big Five variance from the Big Five items had no effect on their ability to predict outcomes, suggesting that there was as much outcome-related personality variance beyond the Big Five as there was in the Big Five, even in items designed to measure the Big Five in the first place. In this case, different intervention strategies may be needed (e.g., focusing on a few key nuances relevant to an outcome) or cost-efficient interventions may even appear less achievable due to the large number of associations that would have to be.

Finally, facet- and nuance-specific observations may help to identify potentially trivial personality-outcome links. For example, Subjective Well-being (SWB) correlates with Neuroticism and Extraversion, but these associations are driven (in opposite directions) by the Depression and Cheerfulness facets, respectively, that already contain items similar to SWB items (Schimmack et al., 2004). Studies using other facets have observed that the Extraversion-SWB association was entirely driven by the Energy Level facet that contains items about being active and enthusiastic, whereas Assertiveness and Sociability facets could play no roles at all. Of course, a correlation between low enthusiasm and low well-being is hardly surprising (Danner et al., 2020; Margolis et al., 2020).

1.5 | A common-method problem

A problem that could plague trait-outcome research is common method overlap: when both traits and outcomes are assessed with the same (generally self-report) method, shared response tendencies could confound them, by inflating them as well as reducing the distinguishability of personality traits and outcomes. For example, McCrae (2015) has estimated that over a third of variance in personality scales (and therefore possibly in many self-report scales assessing outcomes) represents rater-specific response tendencies, many of which would be shared by ratings of both personality traits and outcomes. Indeed, controlling the substantial shared variance among domains (Van der Linden et al., 2010) often results in much weaker outcome correlations (Laidra et al., 2007; Soto, 2021), possibly because some of these confounding factors are removed. To the extent that this applies and the associations due to method overlap are not of substantive interest, there will be correlations between personality domains and outcomes even in the absence of any substantially interesting associations; as a result the true incremental predictive values of facets and nuances may appear smaller than they are because the true null hypothesis is correlations above zero (say, domains and facets respectively explain 10% and 20% of variance in an outcome; when the null hypothesis due to shared method confounds is 5% of variance explained by default, then facets explain more than twice substantively interesting variance in the outcome). It is therefore desirable to remove the common method variance as much as possible. It is of course possible that removing the common variance might only influence the domain-outcome correlations, but not the power that a collection of domains has in predicting outcomes, just as removing all the Big Five variance (and the common method effect) from items may not decrease their combined predictive power (Mõttus & Rozgonjuk, 2021; Seeboth & Mõttus, 2018); this is an empirical question.

1.6 | The present research

To investigate the predictive powers of domain-, facet-, and nuance-level personality traits further, we replicated Seeboth and Mõttus (2018) comparing nuances' associations with a wide range of life outcomes with those of FFM domains and facets. Using a different sample, different personality inventory and different set of outcomes from those of Seeboth and Mõttus (2018), we tested their claim that models including personality nuances as predictors lead to stronger personality-outcome associations than those limited to domains. Constructive replications of this kind can provide stronger evidence for the tested hypotheses than any direct replication could (Lykken, 1968). We tested three hypotheses, preregistered on the Open Science Framework (https://osf.io/x7ebs/):

1. Associations between facets and outcomes tend to be stronger than associations between outcomes and domains, even with the facets residualized for the domains, indicating that facets provide a higher predictive power than domains.
2. Associations between items (indicators of nuances, besides domains and facets) and outcomes tend to be stronger than associations between facets and outcomes, even with items residualized for the facets (and thereby for domains), indicating that nuances provide a higher predictive power than both facets and domains.
3. As observed by Seeboth and Mõttus (2018), the breadth of outcomes (as rated by multiple judges) does not influence the predictive powers of personality traits for them or the degree to which narrower traits out-predict broader ones.

Importantly, these hypotheses pertain to the *collective* (aggregated) associations of domains, facets and nuances with the outcomes (i.e., correlations of *all* domains, *all* facets or *all* items with outcomes) rather than individual trait-outcome associations. Even very small correlations can produce relatively accurate predictions when aggregated (Mõttus et al., 2020). By training and testing the models in seperate sample partitions, we tested personality traits' predictive powers in a way that accounted for both model over-fitting and the fact that more complex models can a priori capture more variance (Stachl et al., 2020; Yarkoni & Westfall, 2017).

Unlike Seeboth and Mõttus (2018), we used the Big Five Inventory–2, which assesses 15 facets, whereas the personality measure used by Seeboth and Mõttus did not assess facet-level traits. But many of our outcomes were psychological constructs themselves (e.g., gratitude, well-being, identity formation), whereas most of the outcomes of Seeboth and Mõttus were more “objective,” directly observable characteristics (e.g., education, BMI) and/or behaviors (e.g., volunteering). Therefore, due to the stronger links between personality traits and a priori somewhat personality-related outcomes, we expected stronger associations than had been observed by Seeboth and Mõttus. Any such greater predictability would not necessarily be substantively interesting, however, as it could merely reflect greater confounding by response tendencies and content overlap.

Besides comparing outcome-predictive powers of domains, facets and items, we also compared the specificities of these predictions, that is, their discriminant validities. For this, we correlated the prediction of each outcome with not just the outcome itself but with other outcomes as well, and compared these cross-predictions for domains, facets and items. For many purposes it may be desirable that models combine the highest possible predictive accuracy for one outcome with low predictive accuracies for others, so that the predictions are maximally outcome-specific. These analyses, not present in Seeboth and Mõttus (2018), were not pre-registered, but we expected items to allow for the highest predictive specificities and domains the lowest.

To estimate the degree to which common method overlap might have contributed personality trait-outcome associations, we ran additional (not pre-registered) analyses in which each item had been residualized for the sum of all other items prior to aggregating them into scales and any further analyses (Neuroticism items were keyed as Emotional Stability beforehand, so that higher scores on all items reflected the socially desirable trait direction). Because BFI-2 scales have the same number of items keyed the same way, their sum represents the shared variance of all traits to the same degree and can be considered an index of general response tendencies, or a general personality factor (GFP), which we deemed substantively uninteresting in this study.

2 | METHODS

2.1 | Participants

We used data from the Life Outcomes of Personality Replication Project (LOOPR Project; Soto, 2019). This project recruited 6,126 US adults, who completed the BFI-2 and a battery of outcome measures adapted from previous research on trait-outcome associations. Due to the large total number of outcomes, the testing materials were split into two surveys (Survey 1 and Survey 2), with each survey including the BFI-2 and a subset of outcome measures. Soto (2019) used quota sampling to ensure that the samples were similarly representative of the US population in sex and ethnicity. Participants were recruited and assessed through the Qualtrics online survey platform and paid approximately $3 per survey. A total of 3,109 participants completed Survey 1, and whilst 3,017 participants completed Survey 2. He assured a minimum of 3,000 participants per survey to provide similarly high statistical power to detect even small trait-outcome associations. The large, basically representative sample, hierarchically structured personality measure, and broad range of outcome measures made the LOOPR Project's data well suited for our purposes.

2.2 | Measures

As noted, personality traits were assessed using the BFI-2 (Soto & John, 2017). This 60-item survey measures the Big Five personality domains, as well as three facets per domain, with 12 items per domain and 4 items per facet. Each item is a short, easy-to-understand phrase (e.g., “Keeps things neat and tidy,” “Tends to be quiet”) with the common item stem “I am someone who…” Individuals rate the extent to which they agree with each item on the typical 5-point Likert scale version. In the combined LOOPR sample, internal consistencies (Cronbach's alphas) for the BFI-2 domain scales were .83 for Extraversion, .79 for Agreeableness, .86 for Conscientiousness, .89 for Negative Emotionality, and .81 for Open-Mindedness. For the BFI-2 facets, the alphas were .77 for Sociability, .53 for Compassion, .78 for Organization, .72 for Anxiety, .67 for Esthetics, .67 for Assertiveness, .70 for Productivity, .78 for Depressiveness, .64 for Curiosity, .64 for Energy, .61 for Trustworthy, .66 for Responsibility, .78 for Volatility and .64 for Creativity.

Outcome constructs and measures for the LOOPR Project were selected from a landmark review of the personality-outcome literature (Ozer & Benet-Martinez, 2006). In total, 83 outcome variables were assessed, using questionnaire measures adapted from the studies they had reviewed. To provide sufficient statistical power for the present research, we excluded outcome measures completed by fewer than 3000 participants, resulting in 53 outcomes for the main analyses. Information about all the LOOPR outcome measures, including their sources, numbers of items, sample sizes, and alphas, is available at https://osf.io/jn7ck/.

3 | OUTCOME BREADTHS VERSUS SPECIFICITIES

All authors of this article, and six other individuals, all of whom had at least master's degrees in, rated each outcome's specificity on a 5 point scale, with 1 being extremely specific and 5 being extremely broad. Risky driving was rated as the most specific outcome ($m = 1.61$) whilst life satisfaction was rated the broadest outcome ($m = 4.82$). The individual-rater intraclass correlation (ICC) of the ratings provided for the 53 outcomes was .27, whilst the ICC of the average ratings across all raters was .80 (Table S1).

4 | ANALYSIS

4.1 | Pre-registered analyses

Our main analyses followed those of Seeboth and Mõttus (2018). To avoid overestimating trait-outcome associations due to over-fitting, the entire dataset was combined into one, before repeatedly splitting it into two random subsamples: training (67%) and validation (33%). In the training samples, we set up elastic net regression models (Zou & Hastie, 2005) for each outcome (with 10-fold cross-validation and shrinkage parameter that minimized cross-validation error within these folds) with either the Big Five domains, their 15 facets, the residuals of the facets (i.e., the facet scores after controlling for the Big Five domains), the 60 items or the residuals of the items (controlling for the Big Five facets and thereby domains) as predictors. We obtained the residuals from linear models wherein each facet/item was regressed on all 5 domains/15 facets, excluding the facet/item being residualized from the domain/facet at the time to avoid residualizing it on itself.

Though ordinary regression can over-fit coefficients (Yarkoni & Westfall, 2017), an elastic net regression instead shrinks coefficients toward 0 and produces more conservative models, which lessen the possibility of over-fitting coefficients. The elastic net regression model treats highly correlated predictors similarly, either including or excluding all of them from model and estimating which combination produces the greatest predictive power (Waldmann et al., 2013; Zou & Hastie, 2005). We then fit the trained models in the validation samples to predict each outcome and correlated these predicted outcome scores with their observed values to calculate prediction accuracies for domains, facets, facet residuals, items, and item residuals, respectively (we squared the correlations to show the percentage of explained variance). We repeated the procedure 100 times in random training-validation sample splits and report the means and standard deviations of these estimates across the splits. We also correlated the outcomes' breadth ratings with the degrees to which the outcomes were predicted by domains, facets and items to estimate how outcomes' breadth can influence predictability.

4.2 | Exploratory analyses

After pre-registering the project, it occurred to us to test the outcome-specificities, or discriminant validities, of the predictions. For this, we correlated the predicted value (from elastic net models using domains, facets or items as predictions) of each outcome with the observed values of other available outcomes. These cross-predictions had to be calculated separately for Surveys 1 and 2 because most participants did not overlap between them, and mutual cross-predictions (correlation between the prediction of outcome x and observed outcome y, and vice versa) were averaged. Because the strongest prediction models for any one outcome (e.g., those based on items) were a priori likely to show the strongest cross-predictions, and the other way around, we normalized the cross-predictions by dividing them by the geometric means of the predictive accuracies for both variables involved in the models specifically trained for them (so that the diagonal of the cross-prediction matrix contained 1s). We calculated the averages and medians of all normalized cross-predictions (squared correlations) for domains, facets and items. Drawing parallels with widely-calculated genetic correlations, Hang and colleagues (2021) call such cross-predictions "personality correlations" between outcomes, whereas Revelle and colleagues (2021) call them "persome correlations".

We also examined the potential effects of the general factor of personality (GFP), likely representing general rating tendencies, on the prediction accuracies, given that this had previously been observed to explain 20%–60% of domain variances (Van der Linden et al., 2010). Besides correlating each outcome with GFP, we also residualized domains for it by regressing each item for GFP (excluding the item being residualized at the time from the GFP) prior to calculating domain scores. As for facets and items, our main analyses that residualized them for domains/facets already controlled GFP.

We repeated both procedures 100 times in random training-validation sample splits.

5 | RESULTS

5.1 | Aggregated associations of item, facet and domain models

Figure 1 and Table S2 (in the Online Supplement; https://osf.io/x7ebs/ the percentages of variance in outcomes for which the three different levels of the personality trait hierarchy accounted.

The 53 outcomes varied considerably in how well personality traits predicted them. For example, Existential Wellbeing, Negative Temperament and Forgiveness were the most accurately predicted outcomes (>45% of variance explained by facets and nuances), while Heart Disease, Substance Abuse and Conventional Interests were the least predictable outcomes (<3% of variance explained at any level).

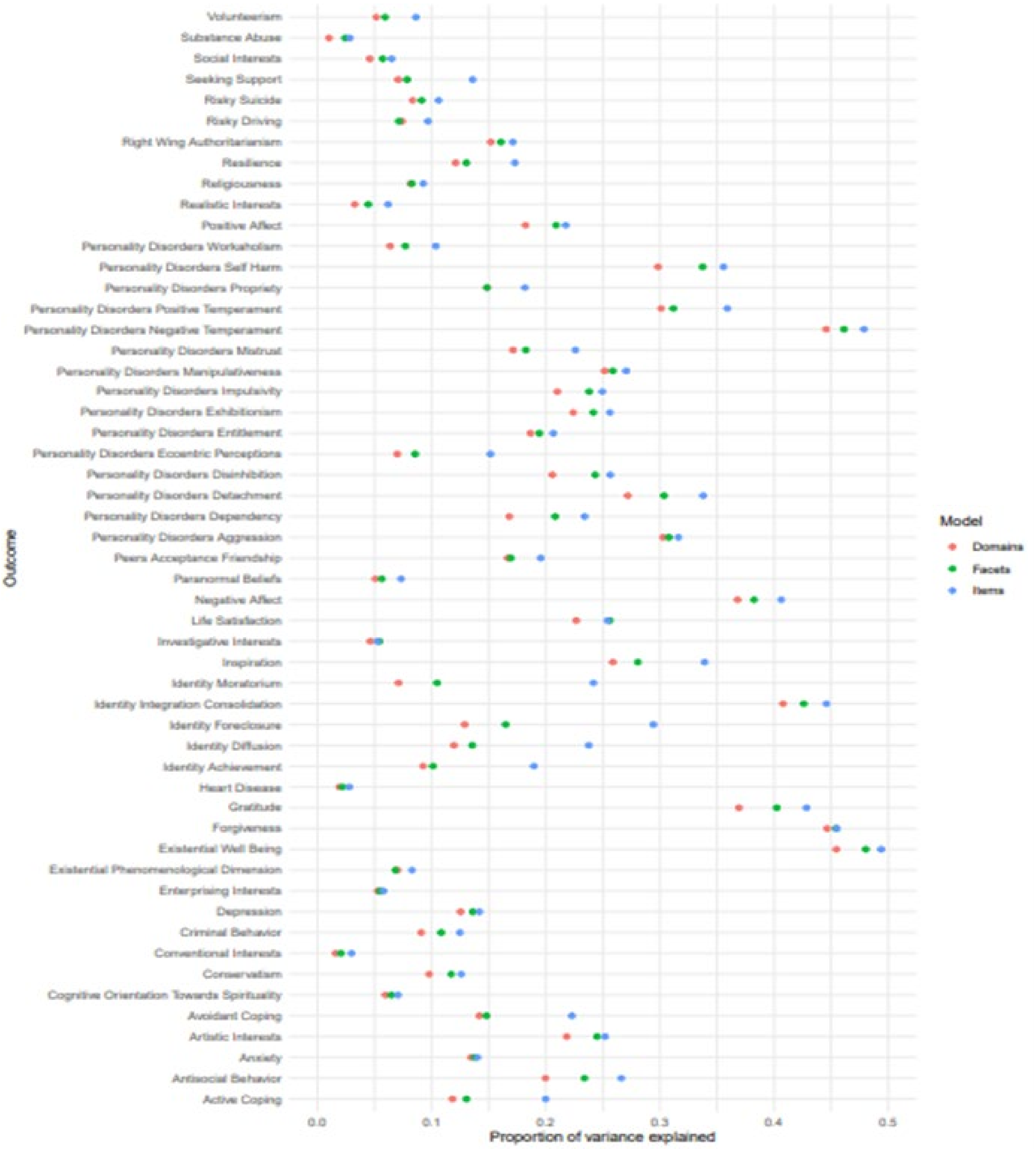

FIGURE 1 Variance accounted for in each outcome by domains, facets, and nuances [Color figure can be viewed at wileyonlinelibrary.com]

On average, domains, facets and items accounted for 16.6%, 18.0% and 20.9% of variance in the outcomes, with respective medians of 13.4%, 14.8%, and 20.1%. Comparing the models at the level of individual outcomes, facets explained on average 14.0% more variance than domains (*Mdn* = 9.5%) and items 23.5% more variance than facets (*Mdn* = 13.7%); items explained on average 42% more variance than domains (*Mdn* = 24.3%). When facets were residualized for domains (and thereby also GFP), they still accounted for 18.0% of variance in the outcomes, on average (*Mdn* = 14.7%), so removing the variance of the Big Five from its facets had virtually no effects on their predictive validities, on average. When items were residualized for facets (and thereby for domains and GFP), they accounted for

20.0% of variance in outcomes, on average (*Mdn* = 19.3%), so residualizing the Big Five items for both the Big Five and its facets had only small effects on their predictive validities. Examining the results for individual outcomes revealed that for 2 outcomes (4%) facet-level traits were the strongest predictors whereas for 51 outcomes (96%) item-level traits (nuances) out-predicted both domains and facets.

These observations supported our first and second hypotheses—that facet-level traits would generally out-predict the Big Five domains and nuances would out-predict facets—and this held regardless of whether the higher-order variance was removed from facets and nuances or not.

5.2 | Outcome breadth

To test whether outcomes' relative breadths were related to their predictabilities from domain-, facet-, and nuance-level traits, we correlated the average breadth scores given by the 11 raters for the outcomes with their average predictabilities at each of the 3 levels. The non-significant correlations were r = .06, .07, and .08 for domains, facets, and nuances, respectively, supporting our third (null) expectation. That is, outcome breadth was not involved in the extent of association with domain-, facet-, or nuance-level personality traits, contradicting Asendorpf et al.'s (2016) proposal. Narrower traits tend to out-predict broader ones for relatively broader or narrower outcomes.

5.3 | Not pre-registered analyses

Item-based predictions had the highest outcome-specificities (discriminant validities), with average absolute cross-predictions of .31 (*Mdn* = .28) in Survey 1 and .24 (*Mdn* = .14) in Survey 2. Domains had the lowest specificities, with average absolute cross-predictions of .45 (*Mdn* = .45) in Survey 1 and .34 (*Mdn* = .21) in Survey 2, and facets were in between, with average cross-predictions of .40 (*Mdn* = .40) in Survey 1 and .28 (*Mdn* = .17) in Survey 2. For clarification: average cross-predictions were higher than average predictive accuracies because they were normalized in relation to the predictive accuracies of models for their own "target" outcomes (see the Analysis section above). Thus, item-level models provided not only the most accurate but also the most outcome-specific predictions.

Partialling the GFP from the BFI-2 items had substantial effects on the psychometric properties of its scales. For examples, the Cronbach's alphas of the domain scales (based on the combined sample of both surveys) decreased from the average of .84 (*Mdn* = .83, range = .79 to .89) to .67 (*Mdn* = .67, range = .60 to .76), and those for the facet scales from the average of .68 (*Mdn* = .67, range = .53 to .78) to .50 (*Mdn* = .50, range = .29 to .68). The correlations among the Big Five domains changed even more dramatically, with the average absolute correlation of .38 (*Mdn* = .39, range = .15 to .52) decreasing to −.24 (*Mdn* = −.23, range = −.52 to −.02). Expectedly, the correlations of the Big Five domains with the outcomes also decreased, with the average absolute correlation across all outcomes and all domains decreasing from .20 (*Mdn* = .17) to .11 (*Mdn* = .09). This decrease was in part because the outcomes that previously overlapped with the Big Five domains most showed strongest decrease in the strength of their correlations after residualizing GFP, indicating that this general variance had played an especially large role in their original associations (Figure S1) and consistent with the observation that the GFP was involved in many outcomes too (Table S4).

Strikingly, however, residualizing items of the Big Five for GFP before aggregating them into the domain scales had almost no effect on how well the domains *collectively* predicted the outcomes, with an average explained variance of 16.5% (*Mdn* = 13.5; see Figure S2 and Table S2), almost identical to the 16.6% before controlling GFP. This suggested that although individual domain-outcome correlations were substantially inflated by their common GFP variance, it was not this common variance that drove their aggregate *predictive* power.

5.4 | Some examples of item-specific associations

Figure S3a,b (Online Supplement; https://osf.io/x7ebs/) illustrate the relations between the items, grouped by facets and domains, and the individual outcomes. We generated three plots for each outcome using the Manhattan function in Psych R package (Revelle et al., 2021): correlations with raw items, with items residualized for GFP, and finally with items residualized for the Big Five domains and their facets (and hence GFP). It is not our purpose here to address each outcome's relations with specific items, but we outline a few general observations. Often removing GFP from items and/or Big Five facets reduced trait-outcome association magnitudes, but in many cases the associations became stronger; sometimes even more items were significantly (corrected using Holm method (Holm, 1979)) associated with outcomes after residualizing than before. Sometimes, associations flipped between positive and negative, varying with whether and how items had been residualized; sometimes associations became more homogeneous within domains and facets, and sometimes the opposite. Overall, (a) most outcomes were still correlated with several, and sometimes even nearly all, items after the items had been residualized for the GFP and the Big Five, and (b) outcomes that were most strongly associated with outcomes before any residualizing were also most strongly correlated with them after residualizing.

So, removing the general variance as well as the Big Five domains from items did change their correlations with outcomes, but often quite unpredictably. This may suggest that while in some cases some of the more broadly defined variance that items were designed to capture "for" the Big Five domains and their facets could have contributed to the items' predictive power, some of it could also have suppressed the models' abilities to indicate the items' full predictive powers, thus resulting in different variance predicting outcomes before and after residualizing items for higher-order variance.

One specific example is how the item-level associations with Identity Foreclosure could be represented and interpreted. (Figure 2). Items from several facets and domains were linked with Identity Foreclosure, although within facets items' correlations varied, often even in direction; suggesting that Identity Foreclosure may be highly "poly-nuanced" in its relations with personality traits. This observation persisted when the models controlled the GFP and the Big Five and their facets. Helping to profile someone high in Identity Foreclosure, among the most closely related items were (from ten different facets): "Is full of energy"; "Is outgoing, sociable"; "Has difficulty imagining things"; "Avoids intellectual, philosophical discussions"; "Has few artistic interests"; "Rarely feels anxious or afraid"; "Feels little sympathy for others"; "Feels secure, comfortable with self"; "Starts arguments with others"; "Assumes the best about people (reverse scored)". Sometimes other items from the same facets had opposite associations with it, such as "Rarely feels excited or eager" (same facet as "Is full of energy") or "Is inventive, finds clever ways to do things" (same facet as "Has difficulty imagining things"). This pattern paints a nuanced picture of someone high in Identity Foreclosure: shrewd yet unimaginative; energetic yet not eager; incurious, not introspective, and somewhat aloof from others.

For criminal behavior, an example of a more specific outcome, the five strongest correlations were with "Sometimes behaves irresponsibly", "Starts arguments with others", "Has a forgiving nature (reverse scored), "Can be cold and uncaring", and "Leaves a mess, doesn't clean up", with correlations from .20 to .27. These items are spread across BFI-2 facets and domains, and it would not be easy to create one overarching trait from them; yet they outline someone who might behave criminally.

6 | DISCUSSION

We compared personality traits' predictive powers at the domain, facet, and nuance levels, finding support for three preregistered hypotheses. First, facets of the Big Five domains had greater predictive power than did the domains themselves, even when the Big Five variance had been removed from the facets. Second, nuances—assessed through individual questionnaire items—had still more predictive power, even when facet and domain variances were removed from the nuances. Third, these patterns held equally for both broad and narrow life outcomes, contradicting Asendorpf et al.'s (2016) hypothesis that outcomes' breadths are related to their predicatabilities from personality traits. Notably, we used a modeling approach—elastic net regression—that penalizes overly complex models (Zou & Hastie, 2005) and trained and validated our models in independent samples. We were therefore quite confident that the greater predictive powers of facet and nuance traits did not simply reflect the larger numbers of nuance and facet traits, and therefore greater risks of capitalizing on chance.

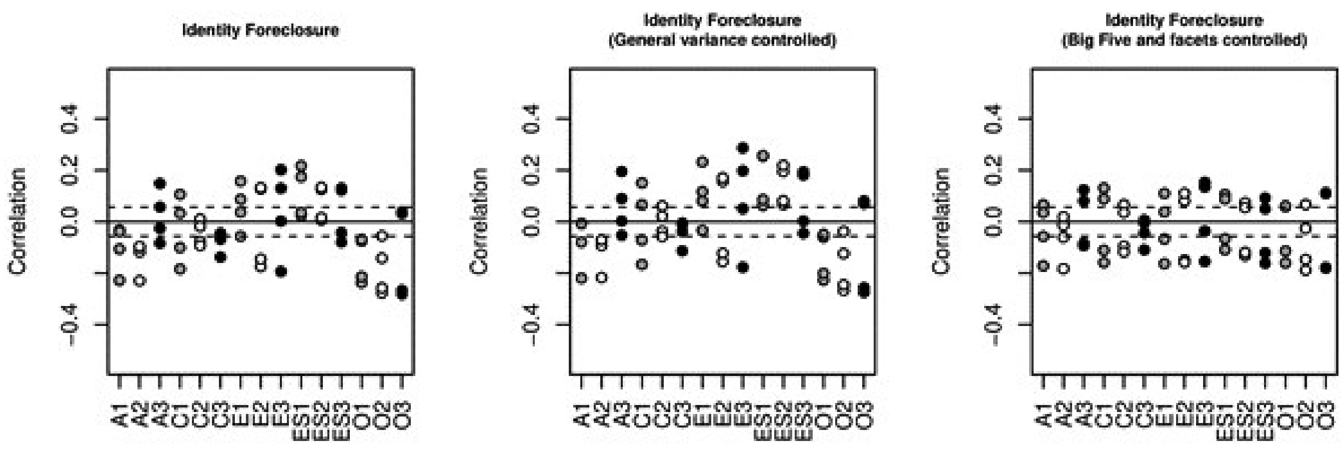

FIGURE 2 Manhattan plots. Each point represents a correlation for a single item with Identity Foreclosure (raw item scores; items after removing the general variance, GFP; items after removing the Big Five and their facets' variance). Items are grouped according to their facets and domains; A = Agreeableness, C = Conscientiousness; E = Extraversion; ES = Emotional Stability (inverse of Neuroticism); O = Openness. Dashed lines indicate the significance threshold after Holm correction of multiple testing

Our pre-registered research replicated the methods and observations by Seeboth and Mõttus (2018), using a different subject sample and investigating different outcomes. However, our models accounted for much more variance; domain models accounted for 16.6% of variance (vs. 4.2%), and nuance-based models explained 20.9% (vs. 5.5%). Plausibly, this was because many of our outcomes were psychological constructs similar to personality traits (or, arguably, personality traits themselves), whereas their's tended to be directly observable behaviors (volunteering) and outcomes (BMI). Additionally, we investigated facet-based models, which accounted for 18% of variance, on average. Our general observation that narrower traits out-predicted broader ones also replicated other studies (Elleman et al., 2020; Revelle et al., 2021; Wessels et al., 2020).

Although we expected that the self-report method overlap between personality traits and outcomes would contribute to their associations, we observed that removing the general self-reported personality trait variance, the so-called GFP, from personality trait ratings had no effects upon the model's predictive powers, despite generally lower correlations between the outcomes and individual personality domains. Removing the GFP as well as the Big Five and its facets' variance from items had varying effects on their outcome correlations. In many cases, these removals did not attenuate the correlations to non-significance and sometimes the item-outcome correlations were stronger after higher-order variance had been removed. Besides indicating that personality trait-outcome associations are often nuanced, this suggested that individual trait-outcome correlations do not indicate much about the combined predictive power of a set of variables (Mõttus et al., 2020): in many cases, several small individual associations can amount to comparatively accurate predictions when combined. This means that descriptive findings (correlations) are different from the predictive power of the variables involved, underscoring the need to not confound the two, as is commonly done.

Finally, the predictions provided by facets and especially nuances were more outcome-specific than those provided by domains, suggesting that nuances had higher discriminant validities too. That is, nuances' incremental predictive accuracy apparently arose because they predicted outcomes' distinctive aspects rather than what the outcomes shared (e.g., their social valence). We expect that outcome-specificity if often a useful feature of representing the involvement of personality in these outcomes.

6.1 | Prediction strength

It may be tempting to dismiss the incremental predictive validities of facets and nuances over domains because the gains have been too small to offset the loss of parsimony. Indeed, we do not argue against using domains to represent personality trait-outcome associations, for sometimes simplicity is paramount, such as for public engagement, and some researchers may a priori prefer parsimon more than others. However, there are strong arguments for not dismissing our observations and their implications outright and for all purposes.

The degrees to which items, indicators of nuances nesodes facets and domains, had stronger outcome associations were admittedly not extensive (e.g., average 21% for items vs. 18% for facets vs. 17% for domains). But, as noted, this pattern was "against the odds" in that BFI-2 items had not been intended to measure nuances: the instrument was carefully designed to measure nothing but the Big Five and their facets. During its development, potential items that appeared to measure something beyond the Big Five and their facets were removed, along with their interstitial contents. Therefore, our observations likely underestimated the potential aggregated predictive power of nuances that could be achieved with broader item sets. In other words, our observations just give a glimpse of what could be achieved, and should motivate further research into how to use nuances to cover personality trait space more comprehensively (Mõttus & Rozgonjuk, 2021; Mõttus et al., 2020).

Another argument for the potential usefulness of nuanced predictions is their higher discriminant validities. Like personality traits, outcomes differ in social desirability and often their predictions may be largely driven by this valence—what is desirable in personality traits predicting what is desirable in outcomes; for evidence, see the correlations between the GFP and outcomes in Table S3. When this boosts apparent predictive power, it is not particularly informative. Facet- and nuance-based models may provide more information, especially when common variance is identified and removed, about distinctive relations between personality traits and outcomes. It may help explain why people with similar Big Few domain scores often experience very different life outcomes, and the other way around.

To the extent that personality-outcome associations are indeed poly-nuanced (which is a feature for us as people, not a "bug"), it may well be that future predictive research will progress by iteratively exploring which combinations of predictor traits maximize the overall and distinctive predictive accuracies of particular outcomes, by systematically trying to eke out every bit of incremental value. Uncovering 3%-4% of predictive power, coupled with greater prediction specificity, would then not be something to be sniffed at, especially if it turns out that the higher limits of outcomes' predictabilities are often relatively models (say 25%). This is a route already established in some scientific fields, such as quantitative genetics where more outcomes are known to be highly polygenetic (e.g., Plomin & von Stumm, 2018).

6.2 | Why?

Why did items, being markers of nuances on top of measuring facets and domains, out-predict the domains and facets? As one possibility, this was due to the *extra* information contained within the nuances, *beyond* the information that the items shared with other items designed to measure their higher-order traits—the "true scores" according to standard psychometric training. For example, the BFI-2 (Soto & John, 2017) contains the Extraversion domain and Sociability facet items "Is outgoing, sociable" and "Is talkative". Actively good listeners tend to be "sociable" but may not be "talkative" themselves, and people who often talk "at" others may often not be seen as very "sociable," for example. As a result, ability and willingness to listen can have unique relevance for a range of outcomes, beyond Sociability and Extraversion more broadly.

But the observation that residualizing items for the Big Five and their facets did not attenuate their predictive accuracy at all is also consistent with a more radical explanation: it is often exactly this *unique* information in items—personality's "details" —that is *mostly* relevant for why people differ in their life outcomes, rather than the more readily interpretable, familiar and easily assessed domain/facet variance that we usually rely on. If so, nuances do not only provide extra details; instead, they provide the whole of the predictive power, giving facets and domains their predictive powers, too. In some cases, broader measures may in fact suppress unique and outcome-relevant information in items, markers of nuances, as evidenced by item-outcome associations often even strengthening once the lower-dimensional variance is removed. The possibility that the unique information in nuances is the actual link between personality traits and outcomes may seem radical and perhaps uncomfortable; it does suggest that much personality research to date may, in some ways, have misrepresented the links.

In any case, it is very plausible that many life outcomes are not directly linked with one or two broad domains but with multitudinous nuances. That is, by sampling from a wide range of items, each relating to specific behaviors and cognitions, more accurate accounts than currently available may be achieved. Properly doing such work requires the bespoke tools for comprehensively assessing personality traits at all levels of the trait hierarchy (Condon et al., 2020).

6.3 | On nuances, again

Re-iterating for emphasis, we do not equate nuances with items per se: items are markers for broader personality traits such as the Big Few domains, but also markers for nuances within and beyond the domains. But many items do represent unique nuances, narrow traits with all the formal trait-properties, even though some are overlapping to the extent that they refer to the same nuance (Condon et al., 2020). Therefore, item-specific research does not imply creating more traits; instead, it means capturing more already-existing trait variances.

However, where single items do capture unique nuances, associations based on them still need to have descriptive—and therefore intuitive—value as well as be comparable across studies. We thus think that future research should strive to operationalize nuances with items that do not refer to specific behaviors in specific situations, but to more situation-general, though specific behavioral traits. This can help researchers avoid the gap between items' contents and the labels that they ascribe to them, which is the causes of the notorious jingle-jangle fallacy plaguing psychometric scales. Items like "Prefers to be alone", "Believes that others have good intentions", or "Feel jealous when hearing about others' successes" can be interpreted as referring to distinct, if correlated, traits that are applicable to many circumstances and to some degree to everyone regardless of their time and location (DeYoung, 2015). In contrast, just what "Enjoys vacationing in Las Vegas" taps is much less clear (and many may never even have considered going there or would have any opportunity to); hence such items should be avoided for measuring nuances.

We believe that future research that uses appropriately measured nuances can accumulate descriptive observations at least as well as broad trait-based research—maybe even more, if the jingle-jangle and social desirability problems with current measures can be better avoided for nuances. Moreover, we also believe that the complexity of nuanced findings can be dealt with by developing novel methods for organizing, summarizing, and visualizing observations and identifying interpretable patterns from them (Mõttus et al., 2020; Revelle et al., 2021). Geneticists and neuroscientists routinely work with very large numbers of observations, having developed tools for extracting meaningful patterns and drawing intelligible conclusions from them. Why could personality researchers not try this too?

6.4 | Breadth

Some have suggested that outcome's and predictors should align: the more specific the outcome, the stronger the associations with more specific traits (Asendorpf et al., 2016). We also tested this, as did Seeboth and Mõttus (2018). Again, our findings were consistent with Seeboth and Mõttus (2018) in that there were *no* relations between outcomes' specificities and the degrees to which they were predicted by either domains, facets or items. In other words, our findings suggest that the differences in prediction strengths between outcomes were due to the types the behaviors they represented, rather than how specific versus broadly aggregated these behaviors appeared to be. Moreover, nuances tend to out-predict broader traits regardless how specific or broad the outcomes are.

6.5 | Limitations & future research

Our research had a number of important strengths, including its large sample size, its hierarchical assessment of personality traits at the domain, facet, and nuance levels, its large and diverse set of life outcomes, and its use of quantitative methods that prevent model over-fitting. However, it also had some important limitations that highlight directions for future research. Perhaps the biggest lay with the measured outcomes that we considered. It can be argued that some of them, such as forgiveness and gratitude, may be better understood as aspects of personality themselves, rather life outcomes. As such, it can be questioned whether our results were representative of relations between items and outcomes, or between items and other would-be (groups of) items. For example, the survey used similar items such as: “I am someone who is temperamental, gets emotional easily” as a measure of Neuroticism (Domain) and “My mood sometimes changes for example, from happy to sad, or vice versa without good reason” as a measure of Negative Temperament (outcome). This is not a problem unique to our particular study, of course, although Seeboth and Mõttus (2018) reported no direct overlaps between their personality test items and outcomes. Once future research can address this concern by applying the methods used in this study, but to different sets of outcomes and inventories, we can get a more detailed picture of the “realism” of nuance relations. For example, we could progress by identifying if nuances are similarly valuable in alternative models, such as HEXACO (Ashton & Lee, 2020) or the MPQ (Tellegen, 2003).

Further to this, investigating nuances beyond the Big Few in models may help to divulge what constitutes a strong predictive nuance. Should two items of rather similar content show similar predictive strengths among various models, this would suggest that the nuance they share has good predictive validity, and should be considered in constructing measurement tools down the line. Conversely, should particular nuances be repeatedly observed to have little predictive validity, they would be best excluded from future inventories. It may even be that outcome-relations will (again) become the primary concern of psychometrics, with inventories developed to reliably investigate individuals' potentials to experience particular outcomes.

As another major limitation, our definition of a nuance here was an item of an inventory specifically designed not to measure nuances. The BFI-2 was carefully crafted to provide the most optimal measurements of the Big Five domains, each defined through three facets (Soto & John, 2017). But the ways in which people think, feel, and behave can be described in many more ways—and apparently these ways uniquely matter for life outcomes. Therefore, future research needs to pursue the most comprehensive yet manageable sets of nuances, perhaps by returning to an Allportian method, and identifying the *many*—and most distinctive—patterns in which individuals behave, rather than just the *major* ways (Big Few domains). By then applying these wider nuance sets to predict outcomes, we can identify the best sets to understand particular outcomes (e.g., Elleman et al., 2020). For example, one set of nuances may be the best predictor of gambling addictions, whilst another set is the best predictor of BMI. One step which could go toward achieving this is publication of item-specific data, even when they are not the focus of interest (Mõttus et al., 2020).

7 | CONCLUSION

In sum, our study extended previous research on personality-outcome associations. In general, facet traits predicted life outcomes more strongly than did broad, domain-level traits like the Big Five, and even narrower nuance traits out-predicted facets. On average, shifting from domain-level to nuance-level traits led to about 25% improvement in prediction accuracies. This suggested that although broad, domain-level traits can provide rough sketches of how personality traits relate with life outcomes, narrower facet- and nuance-level traits are needed to provide more complete pictures. Nuances-based predictions also have higher discriminant validity.

It is exactly the prediction of behavior, some argue, that should be psychology's main goal (e.g., Yarkoni & Westfall, 2017). To do so, we urge looking at nuances, rather than focusing solely on domains, or even facets. Whilst this requires more intricate models, the extra variance they can explain is a beneficial trade-off in some, if not many, circumstances. We stress that we do not argue against the use of domains or facets. Clearly, both do have predictive powers, as this study and others have indicated. Rather, we argue that nuances add another “tool” to the “box” psychology uses to understand how personality contributes to life outcomes, and we have only begun realize this tool's value. As further evidence emerges of these benefits, we call for the most commonly used personality measures to adapt to a hierarchical taxonomy that is actively designed to span the full personality trait space at various breadths that link to outcomes: domains, facets, and nuances, with each level intentionally considered important in its own right. The road to such assessment frameworks has already been mapped by Condon and colleagues (2020), now the onus is on personality researchers to build it!

ACKNOWLEDGMENTS

We would like to thank those who aided with the breadth of outcome ratings for their help.

CONFLICT OF INTEREST

We have no conflicting interests in regard to this article.

ETHIC STATEMENT

All ethical procedures were followed for this study.

AUTHOR CONTRIBUTIONS

Ross David Stewart was responsible for the main write up of the manuscript, with all author's providing feedback and advice at every version. René Mõttus & Ross David Stewart were responsible for the analysis of the data for this study.

ORCID

Ross David Stewart https://orcid.org/0000-0003-1984-3879
René Mõttus https://orcid.org/0000-0002-6403-5224
Christopher John Soto https://orcid.org/0000-0002-2875-8919
Wendy Johnson https://orcid.org/0000-0002-7309-1447

REFERENCES

Achaa-Amankwaa, P., Olaru, G., & Schroeders, U. (2021). Coffee or tea? Examining cross-cultural differences in personality nuances across former colonies of the British Empire. *European Journal of Personality*, *35*(3), 383–397. https://doi.org/10.1177/0890207020962327

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*, *99*, 870–882. https://doi.org/10.1037/a0020963

Anglim, J., & Grant, S. (2016). Predicting psychological and subjective well-being from personality: Incremental prediction from 30 facets over the Big 5. *Journal of Happiness Studies*, *17*(1), 59–80. https://doi.org/10.1007/s10902-014-9583-7

Arumäe, K., Briley, D. A., Colodro-Conde, L., Mortensen, E. L., Jang, K. L., Ando, J., Kandler, K., Sørensen, T. I. A., Dagher, A., Mõttus, R., & Vainik, U. (2020, March 30). Two genetic analyses to elucidate causality between body mass index and personality. *International Journal of Obesity*. https://doi.org/10.1038/s41366-021-00885-4

Asendorpf, J. B., Baumert, A., Schmitt, M., Blum, G., van Bork, R., Rhemtulla, M., & Mõttus, R. (2016). Open peer commentary and author's response. *European Journal of Personality*, *30*, 304–340. https://doi.org/10.1002/per.2060

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, *34*, 492–510. https://doi.org/10.1002/per.2242

Bakker, B., Klemmensen, R., Nørgaard, A., & Schumacher, G. (2016). Stay loyal or exit the party? How Openness to Experience and Extroversion explain vote switching. *Political Psychology*, *37*(3), 419–429. https://www.jstor.org/stable/44132826

Barlett, C., & Anderson, C. (2012). Direct and indirect relations between the Big 5 personality domains and aggressive and violent behavior. *Personality and Individual Differences*, *52*(8), 870–875.

Bleidorn, W., Hill, P. L., Back, M. D., Denissen, J. J. A., Hennecke, M., Hopwood, C. J., Jokela, M., Kandler, C., Lucas, R. E., Luhmann, M., Orth, U., Wagner, J., Wrzus, C., Zimmermann, J., & Roberts, B. (2019). The policy relevance of personality domains. *American Psychologist*, *74*(9), 1056–1067.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, *130*(6), 887–919. https://doi.org/10.1037/0033-2909.130.6.887

Condon, D. M., Wood, D., Mõttus, R., Booth, T., Costantini, G., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Ziegler, M., & Zimmermann, J. (2020). Bottom up construction of a personality taxonomy. *European Journal of Psychological Assessment*, *36*(6), 923–934. https://doi.org/10.1027/1015-5759/a000626

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. University of Illinois Press.

Dam, V., Hjordt, L., Da Cunha-Bang, S., Sestoft, D., Knudsen, G., & Stenbæk, D. (2018). Five-factor personality is associated with aggression and mental distress in violent offenders. *European Neuropsychopharmacology*, *28*, S35–S36. https://doi.org/10.1016/j.euroneuro.2017.12.061

Danner, D., Lechner, C., Soto, C., & John, O. (2020). Modelling the incremental value of personality facets: The domains-incremental facets-acquiescence bifactor showmodel. *European Journal of Personality*, *35*(1), 67–84. https://doi.org/10.1002/per.2268

Deary, I. J., Weiss, A., & Batty, G. D. (2010). Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychological Science in the Public Interest*, *11*(2), 53–79. https://doi.org/10.1177/1529100610387081

DeYoung, C. (2015). Cybernetic Big Five theory. *Journal of Research in Personality*, *56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Dudley, N., Orvis, K., Lebiecki, J., & Cortina, J. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow domains. *Journal of Applied Psychology*, *91*(1), 40–57.

Elleman, L. G., Condon, D. M., Holtzman, N. S., Allen, V. R., & Revelle, W. (2020). Smaller is better: Associations between personality and demographics are improved by examining narrower traits and regions. https://doi.org/10.31234/osf.io/dpmg2

Goetz, K. E., Reeves, M. J., Gagadam, S., Blain, D., Bender, C., Lwin, C., Naik, A., Tumminia, S. J., & Hufnagel, R. B. (2020). Genetic testing for inherited eye conditions in over 6,000 individuals through the eyeGENE network. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, *184*, 828–837. https://doi.org/10.1002/ajmg.c.31843

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037//0022-3514.59.6.1216

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

Graham, E. K., Rutsohn, J. P., Turiano, N. A., Bendayan, R., Batterham, P., Gerstorf, D., Katz, M., Reynolds, C., Schoenhofen, E., Yoneda, T., Bastarache, E., Elleman, L., Zelinski, E. M., Johansson, B., Kuh, D., Barnes, L. L., Bennett, D., Deeg, D., Lipton, R., … Mroczek, D. K. (2017). Personality predicts mortality risk: An integrative analysis of 15 international longitudinal studies. *Journal of Research in Personality*, *70*, 174–186.

Hang, Y., Speyer, L. G., Haring, L., Lee, B., Vainik, U., Luciano, M., & Mõttus, R. (2021, March 31). Establishing a health network of personality profiles for adolescents and emerging adults. https://doi.org/10.31234/osf.io/4qh96

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. http://www.jstor.org/stable/4615733

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level domains in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology*, *74*, 1556–1565. https://doi.org/10.1037/0022-3514.74.6.1556

Jokela, M. (2009). Personality predicts migration within and between U.S. states. *Journal of Research in Personality*, *43*(1), 79–83. https://doi.org/10.1016/j.jrp.2008.09.005

Jones, S. E., Miller, J. D., & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, *39*, 329–337. https://doi.org/10.1016/j.jcrimjus.2011.03.004

Judge, T., Rodell, J., Klinger, R., Simon, L., & Crawford, E. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875–925. https://doi.org/10.1037/a0033901

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*(3), 441–451. https://doi.org/10.1016/j.paid.2006.08.001

Lievens, F., & Johnson, W. (2017). Assessing personality-situation interplay in personnel selection: Toward more integration into personality research. *European Journal of Personality*, *31*(5), 424–440. https://doi.org/10.1002/per.2111

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*(3, Pt.1), 151–159. https://doi.org/10.1037/h0026141

Margolis, S., Stapley, A. L., & Lyubomirsky, S. (2020). The association between Extraversion and well-being is limited to one facet. *Journal of Personality*, *88*, 478–484. https://doi.org/10.1111/jopy.12504

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychological Science*, *27*(5), 715–725. https://doi.org/10.1177/0956797616635200

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

McCrae, R. R., & Costa, P. T. (2010). *NEO inventories professional manual*. Psychological Assessment Resources.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., & Sutin, A. R. (2018). A five-factor theory perspective on causal analysis. *European Journal of Personality*, *32*, 151–166. https://doi.org/10.1002/per.2134

Mõttus, R. (2016). Towards more rigorous personality domain–outcome research. *European Journal of Personality*, *30*, 292–303. https://doi.org/10.1002/per.2041

Mõttus, R., Bates, T. C., Condon, D. M., Mroczek, D., & Revelle, W. (2017). Leveraging a more nuanced view of personality: Narrow characteristics predict and explain variance in life outcomes. https://doi.org/10.31234/osf.io/4q9gv

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

Mõttus, R., Mccrae, R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality*, *52*(C), 47–54. https://doi.org/10.1016/j.jrp.2014.07.005

Mõttus, R., & Rozgonjuk, D. (2021). Development is in the details: Age differences in the Big Five domains, facets, and nuances. *Journal of Personality and Social Psychology*, *120*(4), 1035–1048. https://doi.org/10.1037/pspp0000276

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., Mortensen, E. L., Colodro-Conde, L., & Jang, K. L. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality & Social Psychology*, *117*, e35–e50. https://doi.org/10.1037/pspp0000202

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big few traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311. Let's get going building it!

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421. https://doi.org/10.1146/annurev.psych.57.102904.190127

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews. Genetics*, *19*(3), 148–159. https://doi.org/10.1038/nrg.2017.104

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. . *Personality and Individual Differences*, *169*, 109905. https://doi.org/10.1016/j.paid.2020.109905

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality domains, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet-level analysis. *Personality and Social Psychology Bulletin*, *30*(8), 1062–1075. https://doi.org/10.1177/0146167204264292

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate predictions. *European Journal of Personality*, *32*(3), 186–201. https://doi.org/10.1002/per.2147

Soto, C. J. (2019). How replicable are links between personality domains and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, *30*, 711–727.

Soto, C. J. (2021). Do links between personality and life outcomes generalize? Testing the robustness of trait–outcome associations across gender, age, ethnicity, and analytic approaches. *Social Psychological and Personality Science*, *12*(1), 118–130. http://dx.doi.org/10.1177/1948550619900572

Soto, C. J., & John, O. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*, 117–143. https://doi.org/10.1037/pspp0000096

Soutter, A., Bates, T., & Mõttus, R. (2020). Big Five and HEXACO personality traits, pro-environmental attitudes, and behaviors: A meta-analysis. *Perspectives on Psychological Science*, *15*(4), 913–941. https://doi.org/10.1177/1745691620903019

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

Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology*, *101*, 579–592. https://doi.org/10.1037/a0024286

Sutin, A. R., & Terracciano, A. (2016). Personality domains and body mass index: Modifiers and mechanisms. *Psychology & Health*, *31*(3), 259–275. https://doi.org/10.1080/08870446.2015.1082561

Tellegen, A. (2003). *Multidimensional personality questionnaire–276 (MPQ-276), test booklet*. University of Minnesota Press.

Terracciano, A., Sutin, A. R., McCrae, R. R., Deiana, B., Ferrucci, L., Schlessinger, D., Uda, M., & Costa, P. T. Jr (2009). Facets of personality linked to underweight and overweight. *Psychosomatic Medicine*, *71*(6), 682–689. https://doi.org/10.1097/PSY.0b013e3181a2925b

Vainik, U., Dagher, A., Realo, A., Colodro-Conde, L., Mortensen, E. L., Jang, K., Juko, A., Kandler, C., Sørensen, T. I. A., & Mõttus, R. (2019). Personality-obesity associations are driven by narrow traits: A meta-analysis. *Obesity Reviews*, *20*, 1121–1131. https://doi.org/10.1111/obr.12856

Van der Linden, D., Te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, *44*, 315–327. https://doi.org/10.1016/j.jrp.2010.03.003

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., & Sölkner, J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in Genetics*, *4*, 270. https://doi.org/10.3389/fgene.2013.00270

Wessels, N., Zimmermann, J., & Leising, D. (2020). Who knows best what the next year will hold for you? The validity of direct and personality–based predictions of future life experiences across different perceivers. *European Journal of Personality*, *35*(3), 315–339. https://doi.org/10.1002/per.2293

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. https://doi.org/10.1017/s0140525x20001685

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

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Stewart, R. D., Mõttus, R., Seeboth, A., Soto, C. J., & Johnson, W. (2022). The finer details? The predictability of life outcomes from Big Five domains, facets, and nuances. *Journal of Personality*, 90, 167–182. https://doi.org/10.1111/jopy.12660