7-12-2023

Modeling Variability in Personality Development: Quantifying, Explaining, and Examining the Predictive Utility of Person-Specific Variance Around Personality Trajectories

Amanda Jo Wright

Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological & Brain Sciences

Dissertation Examination Committee:
Joshua J. Jackson, Chair
Michael J. Strube
Sara J. Weston
Diana J. Whalen
Emily C. Willroth

Modeling Variability in Personality Development: Quantifying, Explaining, and Examining the Predictive Utility of Person-Specific Variance Around Personality Trajectories

by

Amanda J. Wright

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2023
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation itself is the result and final continuation of four wonderful years of graduate school. Thus, for not only this project, but accordingly the past four years, there are several people that made it all possible. First, I would like to thank the research and administrative teams behind GSOEP, HILDA, HRS, LISS, and NLSY for providing such great, rich sources of data that have made this project, and many others, possible. Second, I would like to thank my committee members – Michael Strube, Sara Weston, Diana Whalen, and Emily Willroth – that have uniquely been a part of this project and my graduate education, including being the first to teach me advanced analyses, giving me opportunities to grow as an independent researcher, providing valuable advice on my future career, and taking the time to read this paper and offer brilliant suggestions only they could contribute. Third, I would like to thank my friends Jayde Homer, Gabrielle Pfund, and Tabea Springstein for being a wonderful support system and outlet that allowed me to sincerely enjoy my time both in and outside of graduate school. Lastly, I would like to thank my advisor Joshua Jackson for creating the perfect professional and personal environment and for truly being the perfect advisor I could have ever hoped for. I feel as though the dissertation is often one of the most substantial papers many do in graduate school; you have ensured this was but another project I have completed that will be the foundation of (hopefully) a long, joyous career in academia. You have made me a better scientist, colleague, and person; I am forever grateful that I have a lifelong collaborator and mentor in you.

Amanda J. Wright

Washington University in St. Louis

August 2023

ABSTRACT OF THE DISSERTATION

Modeling Variability in Personality Development: Quantifying, Explaining, and Examining the Predictive Utility of Person-Specific Variance Around Personality Trajectories

by

Amanda J. Wright

Doctor of Philosophy in Psychological & Brain Sciences

Washington University in St. Louis, 2023

Professor Joshua J. Jackson, Chair

Decades' worth of research has identified replicable, average patterns of normative personality development across the lifespan. Underlying these findings lie a few assumptions, however, and work seeking to provide incremental knowledge seems to have yet offered a clear path forward. This impasse in progress, in conjunction with studies highlighting the complexity of personality, would suggest these assumptions are perhaps untenable. The current study uses five longitudinal datasets (*N* = 128,345) to examine personality development using mixed effects location scale models. This permits there to be individual differences in within-person residual variability, or sigma, around trajectories – thereby testing if standard models that assume this is homogeneous, unsystematic noise with no implications are appropriate. In doing so, I investigate if there are individual differences in longitudinal within-person variability for trajectories of the Big Five traits; if there are variables associated with this heterogeneity; and if person-level sigma values can predict an outcome, above and beyond the effects of trait levels and changes. Results indicated that, across all models, there were meaningful individual differences in sigma – the magnitude of which was comparable to that of intercepts and slopes. Variables of empirical and theoretical importance in personality development were further associated with this longitudinal within-person variability. Lastly, person-level sigma values proved to have robust predictive utility. Collectively, these results underscore the presence and degree of individual differences in longitudinal within-person variability; their potential for uniquely advancing knowledge; and the necessity of empirically quantifying and theoretically incorporating this individual difference in future personality development research.

.

Chapter 1: Introduction

Decades of work exists showing that although personality is relatively stable across time, it also changes and continues to develop throughout the lifespan (Bleidorn et al., 2022; Roberts et al., 2006; Wright & Jackson, 2023a). Through this research, general trends of personality development, particularly for the Big Five traits (Goldberg, 1990), have been discovered and continue to be refined with recent data (Bleidorn et al., 2022). Though there are some mixed findings within this literature (Graham et al., 2020), there is generally a high agreement on what normative personality stability and development look like. Underlying these many findings of both stability and change lie a few assumptions, however.

First, it is assumed that personality development is appropriately modeled and quantified across time. Typically, changes are modeled with standard multilevel models (MLMs) that are linear, or sometimes quadratic, in form (c.f. Bleidorn et al., 2022; Fraley & Roberts, 2005; Wright & Jackson, 2023b). Although findings from research using these model forms have replicable trends, this does not indicate it is the most appropriate specification. Research in areas beyond normative mean-level personality development (e.g., ipsative consistency, dynamics, life events; Beck & Jackson, 2020; Fleeson, 2001; Wright & Jackson, 2023b, 2023c) would suggest that fitting these simpler trends, which ignores the complexity of personality development, is perhaps inappropriate.

Second, although some individuals adhere to the typical assumed trajectory, others reliably change in different ways, such as those that are cubic, discontinuous, or nonlinear in form (Wright & Jackson, 2023d). That is, there are individual differences in the model forms that best fit individuals' trajectories across time. Due to this heterogeneity, a modeling approach that permits individuals to vary in how well they adhere to their fitted trajectory – which is of a single model form that is imposed on the entire sample – would be most appropriate. However, most models used in general personality development research, including standard MLMs as well as many types of structural equation models (SEMs), instead constrain this degree of person-level fit to be equivalent across the entire sample or restricted in value, consequently misrepresenting patterns of change for many individuals.

Third, it is assumed average trends can be appropriately gleaned from aggregating across many separate individuals. In consideration of the two points above, it seems untenable to expect that a meaningful average can be quantified while imposing the constraints of a) potentially incorrect model specifications and b) identical model forms across individuals that likely differ in the forms of their trajectories. A lot of research has been dedicated to understanding factors underlying individual differences in personality development (Allemand et al., 2007; Bleidorn et al., 2018; Pusch et al., 2019; Schwaba & Bleidorn, 2018). Relative to research on average trends, findings from this work are far less conclusive and many questions remain unanswered. It could be the case that properly accounting for what drives these individual differences is complicated. Alternatively, it may be the case that average personality development was not modeled properly to begin with – rendering its individual differences to be fallible from the start.

Considering personality development may have been incorrectly modeled in a lot of past research, this calls into question whether its findings can be reliably and validly used to understand personality and its associations with other variables. Prediction and explanation are two of the main goals of psychology (Hamaker et al., 2020; Yarkoni & Westfall, 2017). Despite some criticisms of the predictive utility of personality traits (e.g., Mõttus et al., 2020; Salganik et al., 2020), a sizable amount of work exists showing their value for predicting important outcomes (e.g., Beck & Jackson, 2022; Mroczek & Spiro, 2007; Soto, 2021; Wright et al., 2022; Wright & Jackson, 2022a, 2023a). Prediction is never perfect, though, and effect sizes are generally small to medium in magnitude (Götz et al., 2022). Similarly, it is rare that robust explanations are found for the effects of personality (Mõttus et al., 2020; Shmueli, 2010; Yarkoni & Westfall, 2017). Oftentimes, explanatory investigations are far-removed from their theoretical backgrounds and put forth inferences beyond what their data and/or design are capable of. In all fairness, though, human nature is complex and personality processes are highly individualized and multidetermined (Mõttus et al., 2020; Pearl & Mackenzie, 2018; Yarkoni, 2020); it is unsurprising that predictive and explanatory research have some shortcomings. However, instead of accepting this, it is worth exploring ways this research can be optimized. Importantly, more accurately modeling personality development can facilitate improvements as well as inspire new work by providing a comprehensive foundation of descriptive knowledge that can adequately support and motivate all types of future research.

In sum, it is established that personality has both stable and mutable properties, and these mutable properties often reveal themselves in predictable ways across the lifespan. Despite this, work seeking to advance knowledge beyond these average trends and uncover their correlates, causes, and implications seems to have yet provided answers to existing questions or offered a clear path forward for future research. The current study seeks to overcome the limitations of past work and rectify these gaps in research. In doing so, I use five longitudinal panel datasets (N = 128,345) to examine Big Five personality development using mixed effects location scale models (MELSMs; Hedeker et al., 2008). MELSMs relax the standard assumption that the variability of occasion-specific residuals, or sigma, is homogeneous across individuals and unsystematic, meaningless noise. Thus, this analytical approach allows one to quantify individual differences in this within-person variability around personality trajectories – thereby testing if typical models that constrain this to be a fixed value are suitable. To the degree that there is meaningful heterogeneity in sigma, I will then examine if there are variables associated with why some individuals show large variability around their trajectories whereas others have very little variability across their repeated assessments. Lastly, I will test the utility of this individual difference by using person-level sigma values to predict an outcome, thus offering an initial answer of if they have future practical use in personality development research.

1.1 Personality Development Across the Lifespan

Of the multiple ways to conceptualize how personality develops and changes across the lifespan (Roberts et al., 2008), one of the most examined metrics is that of mean-level changes. One meta-analysis by Roberts et al. (2006) found significant changes in all Big Five traits across the lifespan, with the most pronounced change occurring in young adulthood. Furthermore, the Big Five traits differ in their typical patterns of change across time. Typically, there are increases in agreeableness and conscientiousness and declines in neuroticism throughout adulthood (Roberts et al., 2006). This pattern of change reflects the maturity principle (Roberts & Nickel, 2021) and has been found in other longitudinal data (Lucas & Donnellan, 2011; Specht et al., 2011), across cultures (Bleidorn et al., 2013), and with different measures (Graham et al., 2020).

Trait-specific trends for extraversion and openness appear to be less replicable and universal, however. At the broad trait level, Roberts et al. (2006) found no evidence for mean-level changes in extraversion. When examining mean-level trends of extraversion's narrower components of social dominance and vitality, though, opposite trends emerged. Both follow the general pattern of having pronounced changes in young adulthood and less change in middle age, but social dominance continues to increase until middle age, whereas social vitality decreases following the college years (Roberts et al., 2006). Then, openness was initially believed to follow a curvilinear trend across the lifespan, with small increases in adolescence and young adulthood followed by small decreases in old age (Roberts et al., 2006). These trends have replicated in some samples (Lüdtke et al., 2011; Schwaba et al., 2019), though discrepancies sometimes arise when further examining middle and older adulthood. Some studies find increases (Mueller et al., 2016) while others find decreases, particularly in late old age (Schwaba & Bleidorn, 2018).

Recently, a meta-analysis by Bleidorn et al. (2022) examined developmental trends in the Big Five from birth to late older adulthood. Consistent with past meta-analytic work, the most change occurred in young adulthood, with traits expected to change by nearly half a standard deviation during peak rates of change (Bleidorn et al., 2022). There were some exceptions to this average trend, however. Neuroticism had a steady, positive rate of change across the lifespan, with its rate of change only minimally dropping after age 20. Then, conscientiousness deviated the most from the general pattern. In early adolescence, it had small decreases that were followed by its largest changes, occurring around age 20, that were positive in direction. Changes approached zero in middle age and then, around age 70, began to slightly decrease again – similar in size to the decreases observed in adolescence.

Importantly, in addition to these normative age trends for the Big Five, there also exist individual differences in personality development (Mroczek & Spiro, 2003; Wright & Jackson, 2023a). These are unique changes in a person's own average levels of a trait that can differ from the average changes typically observed at the population level. For example, while most people tend to decline on neuroticism as they age (Mroczek & Spiro, 2003), other people could increase while others remain largely unchanged. Furthermore, there is evidence for age-graded individual differences in personality change (Allemand et al., 2007; Bleidorn et al., 2009; Mroczek & Spiro, 2003; Roberts et al., 2001; Robins et al., 2001; Schwaba & Bleidorn, 2018; Terracciano et al., 2005), such that the amount of individual variability varies across different age groups. Generally, the magnitude of individual differences in mean-level trends is greatest in young adulthood, reduces in middle age, and increases again in old age (Schwaba & Bleidorn, 2018).

Furthermore, while personality change can reflect normative maturation (i.e., biological) processes, it can also reflect socialization influences of external factors (Lodi-Smith & Roberts, 2007) such as social roles, life events, and one's daily environment (Bleidorn et al., 2018; Denissen et al., 2019; Roberts, 1997; Robins et al., 2002; Specht et al., 2011; Srivastava et al., 2003). Interactions between people and their environment can both precede and reinforce normative and unique changes in personality (Baltes, 1987; Baltes et al., 1999; Caspi, 1998; Helson et al., 2002; Helson & Stewart, 1994; Roberts et al., 2003). Trait change in the Big Five has been associated with factors such as occupational experiences (Hudson & Roberts, 2016; Jackson et al., 2012), role transitions (Bleidorn et al., 2013), substance use (Wright & Jackson, 2023e), and relationships (Lehnart et al., 2010; Wagner et al., 2015). These experiences are associated with personality change across many ages (Hill et al., 2014; van Aken et al., 2006), indicating personality is generally subject to socialization effects throughout the lifespan. Moreover, the counterpart to socialization effects – selection effects – also play a large role in the extent to which an environment impact one's personality. Acknowledging the unique interplay between selection and socialization effects is not only crucial for obtaining a holistic understanding of personality development, but it also highlights the many opportunities for distinct, individual differences to arise.

1.2 Shortcomings in Typical Past Approaches

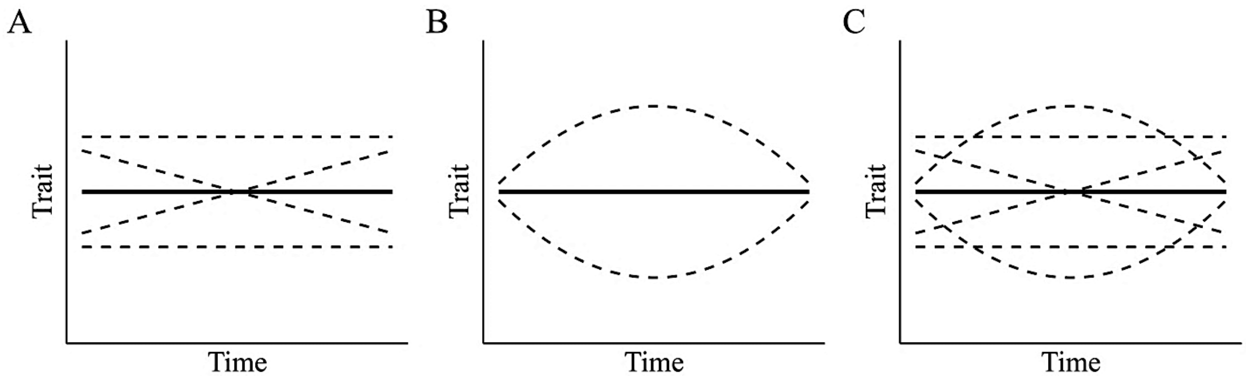
Although past studies (e.g., Lucas & Donnellan, 2011; Lüdtke et al., 2011; Specht et al.,

2011) and meta-analyses (e.g., Roberts et al., 2006; Bleidorn et al., 2022) have generally found replicable mean-level trends for personality development across the lifespan, there appears to be a disconnect with this work and other forms of personality development research. As a result, the implications of these other areas, whose findings often run counter to the assumptions underlying general personality development work, often get ignored. Specifically, past work examining 1) individual differences in trajectories, 2) variability due to environmental factors, 3) individual differences in personality consistency, and 4) personality development using more complex models would suggest the typical way of quantifying average trends is perhaps insufficient.

First, the existence of individual differences in typical mean-level trends has been well-documented in prior research (Allemand et al., 2007; Mroczek & Spiro, 2003; Schwaba & Bleidorn, 2018; Wright & Jackson, 2023a). It is often the case that people's slopes vary in direction and/or magnitude relative to the average modeled trajectory (Figure 1A), which is what random effects in standard MLMs capture. However, people can further vary in the form (e.g., linear, cubic) of their trajectories (Wright & Jackson, 2023d). A constrained linear modeling approach is incapable of taking this into account, though, and will provide an average trajectory of little value. For example, if people have trajectories that are not linear in form, but do not necessarily differ from each other in the average magnitude or direction of their changes (Figure 1B), a linear modeling approach would find no meaningful differences between their trajectories nor relative to the average slope. Furthermore, if people vary in the magnitude, direction, *and* form of their trajectory (Figure 1C), which appears to be the norm rather than the exception (Wright & Jackson, 2023d), then the complexity of these individuals' development is even more poorly captured. This complexity could be captured, though. Specifically, the degree to which people's actual datapoints vary around their estimated trajectories can be quantified by allowing that residual variability to differ across individuals. That is, the homogeneity of variance assumption can be relaxed. Typical modeling approaches assume this value is fixed and equivalent across all individuals – an assumption that may not be justifiable.

Figure 1

Example Personality Trajectories Depicting Individual Differences in Change and Model Forms



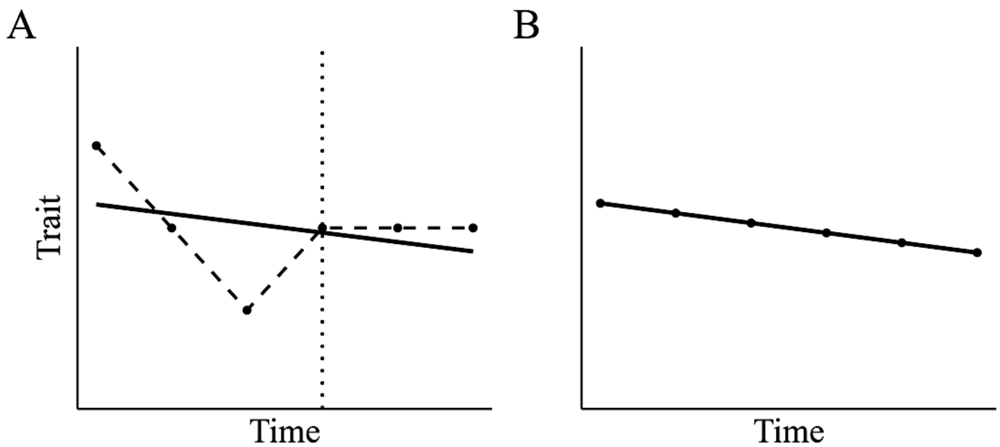
Note. Dashed lines represent the person-level trajectories whereas solid lines represent the modeled sample-level trajectories.

Second, the factors underlying these individual differences have implications. A sizable amount of research has examined the associations that external factors have with long-term personality change and short-term dynamics, namely major life events (Denissen et al., 2019; Schwaba & Bleidorn, 2019; Specht et al., 2011; Wright & Jackson, 2023c) and contextual factors at smaller time scales (Kuper et al., 2022), respectively. In research focused on long-term change, an individual's trajectory is sometimes permitted to vary across time as a function of the event, often in the form of piecewise models (Denissen et al., 2019; Wright & Jackson, 2023e, 2023c). Unfortunately, research on life events is typically inconclusive and offers up null, small, or contradictory empirical and theoretical findings (Denissen et al., 2019; van Scheppingen et al., 2016; Wright & Jackson, 2023c). However, despite complicated conclusions for average effects, a recurring finding is that there are quantifiable individual differences around these effects (Denissen et al., 2019; Wright & Jackson, 2023c). This is important, as it indicates that not only are there meaningful person-level differences in personality trajectories across the chosen time metric, but they additionally exist around distinct, non-time-structured periods (i.e., event boundaries) within a study's duration. If there are indeed significant, inconstant changes in individuals' slopes, then, when examining general personality trends, constraining the slopes to be linear results in an uneven spread of residuals around their trajectories due to these event-induced deviations not being taken into account. A linear model is ignorant to these deviations as it assumes they do not occur and thus constrains the spread of residuals around trajectories to be homogeneous. Ultimately, this provides an inaccurate, simplified representation of how personality is changing across time.

For example, prior to getting married, a person may, on average, decline in a trait; begin to increase in the time immediately leading up to the marriage; and then plateau following the event (Figure 2A). It can be demonstrated how an average linear line of best fit for this individual a) oversimplifies their trajectory and b) differs in the degree of the spread of its residuals relative to the trajectory of someone who instead maintained steady changes in this trait across time (Figure 2B). Instead of allowing the model to quantify this variability at the within-person level, it is then likely, and improperly, attributed to between-person variance or meaningless error. Given it has been shown there are individual differences in the associations that external factors have with personality development (e.g., Denissen et al., 2019; Wright & Jackson, 2023c), and that it is highly improbable for someone to avoid encountering any factor in their environment that could be associated with personality change, the default should be to always account for this heterogeneity. Importantly, constraining sigma to be homogeneous across individuals not only harms the precision of model-estimated effects (Hamel et al., 2012; Jahng & Wood, 2017; Leckie et al., 2014), but it further ignores valuable person-level variance that could account for unexplained heterogeneity in personality development, have associations with theoretically relevant variables, and meaningfully predict future life outcomes.

Figure 2

Example Personality Trajectories for Two Individuals in the Context of a Life Event

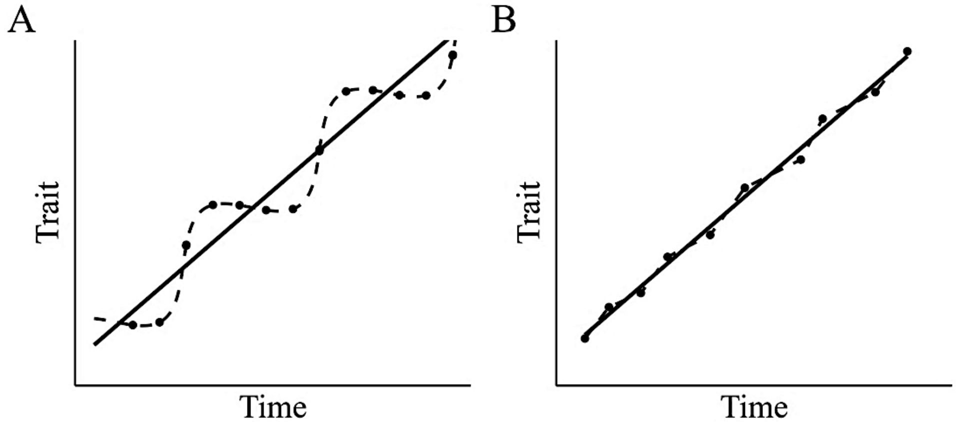


Note. Dashed lines represent each person's actual trajectory whereas solid lines represent their modeled linear trajectory. The vertical dotted line in panel A marks the life event occurrence.

Third, past research has found stable individual differences in levels of Big Five personality profile consistency across time (Wright & Jackson, 2023b). That is, whereas some people are stably consistent in their personality profiles, such that they show minimal changes in their pattern of responding for all trait indicators across time, others are stably *inconsistent*, such that they show greater changes. When considering these patterns through the lens of a single trait, an individual that fluctuates between increasing and decreasing in their trajectory (Figure 3A) will differ from an individual that consistently increases (Figure 3B). Person A not only has greater within-person variability around their line of best fit, but they further had short-term perturbations that accumulated into meaningful increases over time. This is in contrast with Person B who has very little spread around their trajectory *and* monotonically increases across time. Typical models would not be able to differentiate between these trajectories. Considering there are likely important person-level or environmental factors associated with why individuals exhibit these distinct trajectories, not only are these patterns not even identified, but the mechanisms underlying this heterogeneity in personality change further cannot be investigated.

Figure 3

Example Personality Trajectories for Two Individuals That Differ in Profile Consistency

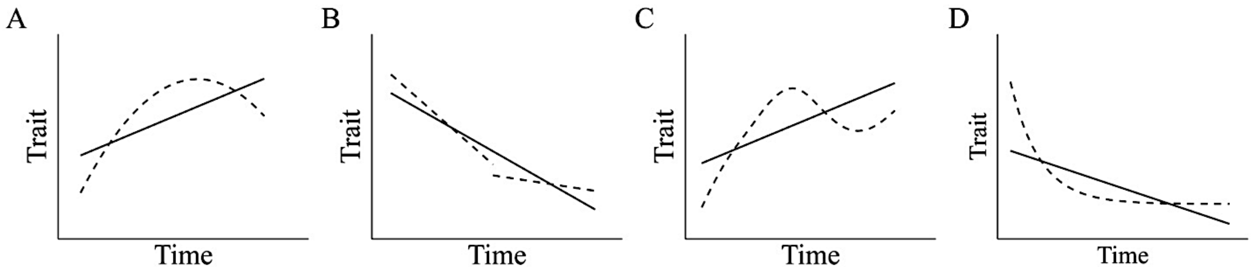


Note. Dashed lines represent each person's actual trajectory whereas solid lines represent their modeled linear trajectory.

Fourth, past research leveraging more complex models would suggest there are important nuances to be captured that get ignored when using typical approaches. Sometimes models with forms beyond simple linearity are fit, such as curvilinear (Figure 4A), discontinuous (Figure 4B), spline (Figure 4C), and asymptotic models (Figure 4D; Bleidorn et al., 2022; Fraley & Roberts, 2005; Wright & Jackson, 2023b). In research examining general personality development trends, curvilinear trajectories, specifically quadratic, are by far the most common when forms beyond linear are fit (e.g., Graham et al., 2020; Lucas & Donnellan, 2011; Roberts et al., 2006). However, while past work using these model forms did allow personality to be modeled beyond simple linearity, the requirements that all individuals fit a single model form and have equivalent residual variability around their trajectories were still imposed. This constraint is not tenable if people do not all share that single, and at times somewhat niche, model form. Moreover, even in data with this average pattern, individuals' trajectories likely vary in their degree and presence of non(linearity), and thus should still be permitted to vary in how well they adhere to that form.

Figure 4

Example Trajectories for Curvilinear, Discontinuous, Spline, and Asymptotic Models



Note. Dashed lines represent each person's actual trajectory whereas solid lines represent their modeled linear trajectory.

Ultimately, in any of these models, individual differences in quality of fit are still not quantified. This makes it impossible to determine if and to what degree there is variability in individual-level adherence to a model form. Importantly, other modeling approaches, typically those used in an SEM framework, often have fewer assumptions and overcome some of the frequent issues seen with linear models. Although progress has been made using these other models, their treatment of individual differences in longitudinal within-person variability is still less than ideal and they possess other limitations that make them ill-suited for comprehensive investigations of sample- and individual-level personality development.

1.3 Approaches That Have Gone Beyond Constrained Linearity

Despite some of the limitations inherent in traditional modeling approaches, past research employing these models has undeniably led to advancements in knowledge about personality development across the lifespan, both normatively and in the context of external factors. However, it is also worth examining other modeling strategies that go beyond the typical constrained linearity approach, as they can provide answers to questions that simpler models cannot address due, in part, to the assumptions inherent to those models. In particular, these more complex analytical approaches can be advantageous because they are predicated on the fact that deviations from some model-implied, predicted score exist; these deviations differ across people; and they need not be modeled with a structured form.

Specifically, models such as the random intercept cross-lagged panel model (RI-CLPM; Hamaker et al., 2015); latent growth curve model with structured residuals (LGCM-SR; Curran et al., 2014); autoregressive latent trajectory (ALT) model (Bollen & Curran, 2004; Bollen & Zimmer, 2010); and stable trait, autoregressive trait, and state (STARTS; Kenny & Zautra, 2001) model have taken different but complementary approaches at modeling additional sources of variance. While research using these models has provided novel and valuable information about personality and its associations with other variables, there is still something left to be desired that these models collectively cannot answer regarding individual-level personality development: insight into who is more or less variable around their personality trajectories across time.

Among the more commonly used analytical approaches that go beyond simple linear trajectories is the RI-CLPM (Hamaker et al., 2015). The RI-CLPM is an extension of the traditional CLPM, such that it allows one to additionally control for stable trait factors, thus often providing less distorted estimates of cross-lagged effects (Lüdtke & Robitzsch, 2022). Although it has benefits over the CLPM, its issues have also been well-documented (Andersen, 2022; Asparouhov & Muthén, 2022; Lucas, 2023; Lüdtke & Robitzsch, 2022). Additionally, even though an average trajectory of any form can be fit, individual differences in change cannot be estimated as only a random intercept is included in the model (Usami et al., 2019). Furthermore, the occasion-specific variance is permitted to vary across waves, but its precise estimation is contingent upon the accurate estimation of other effects in the model that are known to be faulty at times (i.e., cross-lagged effects). Moreover, the occasion-specific variances are uncorrelated with each other across time, which implies they are simply a residual artifact of all effects at a given time point, not a meaningful, person-level individual difference. Consequently, the model further does not actually provide information about which individuals have more or less of this variance. Thus, this modeling strategy is less than ideal for rigorously examining multiple aspects of variability in individual-level personality development.

Then, one model that can serve as an extension of the RI-CLPM is the LGCM-SR. Specifically, a RI-CLPM with a random slope is similar to a LGCM-SR (Curran et al., 2014). The focus of this model is often on within-person regressions that use time-adjacent residuals, or the deviations from the modeled trajectory, to quantify how these residuals are related across time (Curran et al., 2014). Thus, it provides information regarding autoregressive relationships, similar to the RI-CLPM, but does so while additionally accounting for between-person effects of slopes. Importantly, the person- and time-specific residual from the within-person regressions is now the only "true" residual, as it is the portion of the original residual that remains following the partitioning of the average autoregressive effect. Subsequently, the residual is a residualized version of itself. Moreover, excluding the first time point, the residual variance is fixed in value across time and not considered substantively meaningful. Thus, much like the RI-CLPM (Andersen, 2022; Asparouhov & Muthén, 2022), the LGCM-SR relies on residualized estimates and provides no information about individual differences in residual variability.

Next, a model that is further distinguished from the LGCM-SR is the ALT model (Bollen & Curran, 2004; Bollen & Zimmer, 2010). There are some instances in which the LGCM-SR and ALT model are equivalent (Hamaker, 2005), but, generally, the ALT model quantifies the autoregressive relationship between an individual's repeated measures across time whereas the LGCM-SR does this with the residuals. Although the ALT model permits the occasion-specific residuals to vary across individuals, they are not meaningfully examined nor associated with other variables in the model. Indeed, an assumption of the model is that the residual variance at a given time point is uncorrelated with other variables *and* unrelated to its later values. Also, a complication with the ALT model is that the intercepts and slopes are “accumulating factors” (Usami et al., 2019). This means the precise combination of intercepts, slopes, initial average levels, autoregressive effects, and cross-lagged effects for all variables is necessary to obtain the average and individual-level trajectories for a construct (Orth et al., 2021). Thus, if one were modeling a personality trait and some outcome, the intercepts and slopes for the trait could not be interpreted in nor are they quantified in the same manner as a traditional growth curve model. Rather, their interpretations are contingent upon relative comparisons to trait values for other individuals, the accumulating intercept and slope factors for the outcome are controlled for in their estimation, and between- and within-person variance is not fully disentangled (Murayama, 2022; Orth et al., 2021; Usami et al., 2019). Overall, qualities such as these limit the utility of the ALT model for examining both sample- and individual-level personality development.

Lastly, the STARTS model decomposes variance into three components: a stable, time-invariant quantity; a somewhat stable time-varying, autoregressive quantity; and an occasion-specific state quantity (Kenny & Zautra, 2001; Lüdtke et al., 2018). Although the model decomposes multiple sources of variance, it has some limitations. First, there is a stationarity assumption, meaning the variances, correlations, and stability coefficients are fixed values across time (Kenny & Zautra, 1995). If there is any degree of heterogeneity, this assumption is untenable and immediately precludes the ability to quantify individual differences. Second, the state component – the variance of which is the term most comparable to sigma – has little utility. This value captures one’s time-specific deviation from the sum of the sample average, their stable value, and their autoregressive value at that time point. However, it does not differentiate between error and time-specific effects and is uncorrelated with all other factors at any other time point, meaning it cannot be used to examine associations with other components. Third, there is a disturbance term that also varies across time, but it is a function of the autoregressive trait stability and variance. In light of the stationarity assumption, not only is this residual variance a fixed value across time, but it further is not freely estimated. Fourth, individual trends are not estimated nor is a linear change component, which restricts any estimate of systematic change to take the form of an average autoregressive effect. Ultimately, this model is not ideal for examining multiple individual differences in personality development.

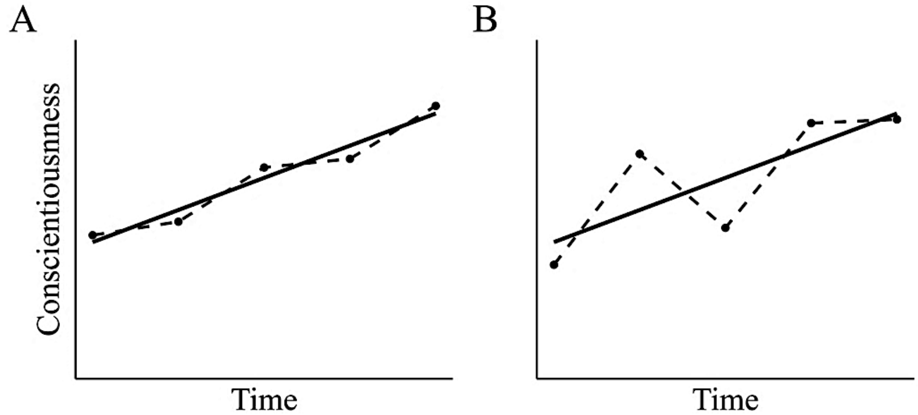
1.4 Mixed Effects Location Scale Models

Although the models in the previous section have some limitations in the context of being able to comprehensively investigate individual-level personality development, they do provide valuable ways of disentangling variance that are less constrained compared to traditional models such as MLMs. However, there is a modeling approach that can rectify many of the issues present in all other models, is free of potentially untenable assumptions, and can be used to more fully capture the complexity of personality development. Specifically, that model is the MELSM (Hedeker et al., 2008). In this model, variance components for between- and within-person effects are separately – and thoroughly – modeled, with only the LGCM-SR offering a similar level of variance decomposition. Moreover, in addition to quantifying variability in intercepts and slopes, the homogeneity of variance assumption is relaxed in MELSMs. Thus, individual differences in the within-person residual variances are estimated, and individuals with more or less of this variability can be identified. Additionally, covariance parameters amongst the between- and within-person random effects can be estimated, which offer unique information about individual-level development and can also account for less-than-desirable measurement properties (e.g., ceiling or floor effects; Hedeker et al., 2008). Furthermore, covariates can be used to examine associations with these variances (Hedeker et al., 2008). Lastly, the person-level sigma values can be extracted and used in further analyses, much like intercepts or slopes.

To demonstrate a scenario in which a MELSM would be useful, an example with conscientiousness and the outcome of health status is described. Generally, conscientiousness tends to slightly increase across adulthood (Bleidorn et al., 2022), and higher mean levels (Bogg & Roberts, 2004; Jackson et al., 2015; Wright et al., 2022) as well as increases in this trait are both associated with better health outcomes (Takahashi et al., 2013; Turiano, Pitzer, et al., 2012; Wright & Jackson, 2023a). Now, suppose two individuals have equivalent intercepts (b_0 = 2.97) and slopes (b_1 = 0.27) for their trajectories of conscientiousness across time, with both having slightly higher-than-average levels and slight increases. Person A, who has consistent increases and whose data points adhere closely to their line of best fit (Figure 5A), aptly also has higher-than-average health status (HS_A = 8). However, Person B, who has steady increases according to their line of best fit (Figure 5B), has average health status (HS_B = 5), which is lower than expected. Importantly, despite Person A and B having equivalent intercepts and slopes, their raw trajectories differ. Using traditional models, a researcher would be ignorant as to potential reasons why Person B has a lower health status, especially given that their line of best fit is equivalent to that of Person A. When using MELSMs, though, one can quantify that Person B has a relatively more haphazard way of responding across time, such that they routinely evidence larger deviations from their line of best fit. More specifically, they have a larger sigma value (σ_B = 0.38) compared to Person A (σ_A = 0.11). Furthermore, this individual difference can further be associated with other variables such as age, cognitive ability, or life events. The answer to why Person B has a lower health status can be thoroughly investigated using this modeling approach.

Figure 5

Example Trajectories of Conscientiousness for Two Individuals That Differ in Their Degree of Within-Person Variability



Note. Dashed lines represent each person's actual trajectory whereas solid lines represent their modeled linear trajectory.

As seen in the scenario above, the MELSM allows one to answer certain questions regarding the nature and implications of an individual's personality development at a more nuanced level compared to other models. Although the above example was specifically chosen to showcase a strength of the MELSM, in personality development research, this modeling approach should generally be preferable to all other models described previously. This is broadly true as it provides more precise parameter estimates and comprehensively disentangles variance. However, it is especially true when one's goal is to thoroughly examine sample-level personality development, individual differences in personality development, and/or the associations of these individual differences with other variables. More specifically, other models are simply not set up to answer the same type nor the breadth of questions that MELSMs are well-suited to investigate.

First, for the RI-CLPM, it can primarily be used to examine how an individual's value for some outcome changes in light of an individual deviating from their typical level of a predictor. Researchers are restricted to examining the effects of within-person deviations in levels of the constructs as individual differences in how constructs change across time, much less individual differences in within-person variability around these changes, are not and cannot be obtained. Second, the LGCM-SR can examine how changes in one's level of a predictor, that differ from their expected level given their estimated trajectory, are associated with changes in an outcome. However, this deviation is conceptualized as a time-specific change in one's trajectory. Thus, conclusions are drawn only from this deviation at a single time point, ignoring how this might be part of a broader pattern of a person's general degree of within-person variability. Moreover, considering autoregressive effects of the residuals are also examined, this further complicates the ability to interpret sigma in these models. Third, the ALT model can answer questions such as how rank-order changes in a predictor are associated with rank-order changes in an outcome. The accumulating intercept and slope of the outcome are controlled for in these effects, though, resulting in the estimated between- and within-person variances – which do not include within-person variability around trajectories in the first place – not even being fully separated (Usami et al., 2019). Fourth, a predictor's association with an outcome in the STARTS model has the same interpretation as in the RI-CLPM, while also having additional constraints (Orth et al., 2021). Lastly, due to their ability to estimate random effects for intercepts and slopes, standard MLMs can examine between- and within-person effects for levels and changes in a construct as well as their associations with an outcome. Thoughtful model specifications can result in a multitude of interesting effects; however, the within-person residual variability is assumed to be and treated as unsystematic noise. Accordingly, sigma is a constant, singular value across time and people in these models, ultimately limiting their utility for investigations beyond effects of level and slope.

Overall, while the above models have their advantages and can each be the optimal analytical framework for certain research questions, when the topics of investigation are multiple aspects of variability in sample- *and* individual-level personality development, they are less than ideal. The nuanced view of heterogeneity in intercepts, slopes, and within-person residual variability around trajectories offered by MELSMs enables them to provide novel information about how personality develops, which factors are associated with individuals' unique trajectories, and the utility of these person-level sources of variance for personality psychology.

1.5 Current Study

Given the decades of personality development research that has resulted in a cumulative and replicable body of literature on how mean levels of the Big Five traits tend to change across the adult lifespan, it may seem unnecessary to probe these findings. However, research seeking to advance knowledge beyond these average trends, such as what drives or moderates individual differences in change, how best to capture development at the individual level, and why static levels and changes in traits seem to not tell the full story of how personality impacts people's lives, has yet to offer a clear and conclusive path forward. A potential solution is to examine an understudied component of the culmination of this research: examining *if* all individuals can be appropriately modeled how has been previously done. By extending prior research and obtaining a more nuanced view of how personality develops at the individual level, theories can be refined, additional parameters of interest can be quantified and contextualized, and a new individual difference metric can be used to examine meaningful associations with important life outcomes.

The advantages of MELSMs suggest they may offer a novel perspective on personality development. Specifically, a precise view of how well people adhere not only to the average personality trajectory across time, but further how well they adhere to their own line of best fit is provided. Within-person variability around trajectories can be wholly modeled and quantified, thus no longer considering it to be homogeneous across individuals or random noise with no implications. Furthermore, one can examine if there are factors associated with why some people remain consistent around their trajectory whereas others have greater within-person variability. Thus, if there are meaningful individual differences in this person-specific variability around personality trajectories, perhaps answers to existing questions can emerge by using previously examined variables of interest to see if they can explain why people differ in their personality development. Moreover, this individual difference can be used in valuable applications, such as the prediction of important outcomes or explanation of personality processes, which can further highlight the circumstances in which it is important to model this person-level variability.

The current study seeks to rectify shortcomings and gaps in past research and offer an attempt at finding solutions to key, unanswered problems. In doing so, I use five longitudinal panel datasets (N = 128,345) to examine personality development using MELSMs. Importantly, this allows individual differences in longitudinal within-person residual variability around trajectories, or sigma, to be quantified. To the degree that there is meaningful heterogeneity in this person-level variability, I will then test if variables that are routinely associated with other aspects of personality development can explain why some individuals may show large variability around their trajectories whereas others adhere quite closely to their line of best fit across time. Lastly, I will examine the predictive utility of this longitudinal within-person variability, in addition to static levels and changes in personality, to determine if it offers novel insight on an empirically robust personality-outcome association.

Chapter 2: Methods

2.1 Participants

In this paper, I use data from N = 128,345 participants from five longitudinal panel datasets. To be included in the study, a participant needed at least one wave of Big Five data. See Table 1 for descriptive information per dataset as well as for all datasets combined. The Institutional Review Board (IRB) at Washington University in St. Louis deemed this project exempt from IRB approval because it involves accessing publicly available datasets and thus does not meet federal definitions under the jurisdiction of an IRB (IRB ID#: 202302142).

Table 1
Individual Dataset and Overall Study Descriptive Information

	Dataset					
Variable	GSOEP	HILDA	HRS	LISS	NLSY	All
N	55,584	24,685	23,533	16,240	8,303	128,345
Age (M)	44.05	38.99	64.85	44.32	20.82	45.42
Age (SD)	17.21	18.68	11.14	17.95	4.91	19.44
Age (Range)	17-98	15-100	19-105	16-100	14-41	14-105
% Female	52%	52%	58%	54%	49%	53%
# waves Big Five (M)	2.11	2.80	2.21	3.81	2.70	2.51
# waves Big Five (SD)	1.29	1.52	1.08	3.01	0.82	1.70
# waves Big Five (Range)	1-5	1-5	1-5	1-11	1-5	1-11

Note. N = sample size. M = mean. SD = standard deviation. Age is the initial age in each dataset.

2.1.1 German Socioeconomic Panel (GSOEP) Study

GSOEP (Socio-Economic Panel, 2020) is an ongoing longitudinal study conducted by the German Institute of Economic Research (DIW Berlin). Data on individuals in more than 11,000 German households are collected and are available by application at https://www.diw.de/soep. Data collection began in 1984 and continues annually, with the latest release in 2021. Data from 2005-2020 were used in the current study. Through 2005-2017, the Big Five were assessed every four years. An additional wave occurred in 2019. Other variables are typically assessed annually.

2.1.2 Household Income and Labour Dynamics in Australia (HILDA) Study

HILDA (Watson & Wooden, 2012) is an ongoing longitudinal study that collects data on more than 17,000 individuals in Australian households. Data are freely available by application at https://melbourneinstitute.unimelb.edu.au/hilda/for-data-users. Data collection began in 2001 and continues annually, with the latest release in 2021. Data from 2005-2021 were used in the current study. The Big Five are assessed every four years, whereas questions regarding other variables are typically assessed annually.

2.1.3 Health and Retirement (HRS) Study

HRS (Juster & Suzman, 1995) is an ongoing longitudinal study of more than 35,000 individuals in the United States. Data are freely available at https://hrs.isr.umich.edu. Data collection began in 1992 and continues biennially, with the latest release in 2020. Data from 2006-2020 were used in the current study. Generally, the Big Five are assessed every four years for an individual, although a small number (n = 43) sometimes have an assessment gap of only two years. Questions for other variables are typically assessed biennially for all participants.

2.1.4 Longitudinal Studies for the Social Sciences (LISS)

LISS (Scherpenzeel & Das, 2010) is an ongoing longitudinal study of Dutch-speaking individuals from 5,000 households in the Netherlands. Data are freely available through application at https://statements.centerdata.nl/liss-panel-data-statement. Data collection began in 2007 and continues annually, with the latest release in 2022. Data from 2008-2022 were used in the current study and questions for all variables were assessed annually.

2.1.5 National Longitudinal Survey of Youth (NLSY) – Children & Young Adults

NLSY (Bureau of Labor Statistics, 2020) is an ongoing longitudinal study of the children of participants from the original National Longitudinal Survey of Youth 1979 (NLSY79). Data are freely available at https://www.nlsinfo.org/investigator/pages/login. NLSY79 consists of data collected on more than 12,500 individuals in the United States since 1979. Data collection began in 1986 and continues biennially, with the latest release in 2020. Data from 2006-2018 were used in the current study and questions for all variables were assessed every two years.

2.2 Measures

2.2.1 Big Five

The primary variables of interest in this study are the Big Five personality traits (Goldberg, 1990). Measures and items varied across datasets (see File S1) and all psychometric information can be found in Table S1. Internal consistency estimates were calculated using the psych package (Revelle, 2021) in R. Due to the different scales of measurement across datasets, all traits were transformed into Percentages of the Maximum Possible (POMP) score (Cohen et al., 1999) so parameters could be compared. In contrast with standardization, POMP preserves the original sample variance and instead relies on the ratio of the differences between an observed score and the minimum possible score and the maximum and minimum possible scores of the scale, or $POMP = \frac{observed - min}{max - min} * 10$. In order to aid in interpretation with other variables and avoid convergence issues, scores were transformed on a 0-10 scale as opposed to 0-100.

2.2.2 Covariates/Moderators

Seven domains of covariates/moderators were included: gender, age, cognitive ability, education level, income, personality traits, and life events.

Gender. Gender was a dichotomous variable, coded such that 0 = male and 1 = female.

Age. In each dataset, a person's age at their first wave of data was centered around the average age at that time point and then scaled (i.e., standardized). Thus, this variable quantified how far a participant's initial age deviated from their dataset's average initial age in standard deviation units. A squared age variable was also calculated.

Cognitive Ability. The availability and content of cognitive ability measures varied across datasets, but when present, cognitive ability was composited across available measures (quantified via standardized scores) across an individual's waves of data. This aggregate value was then also standardized within each dataset. Cognitive ability was not assessed in LISS.

Education Level. Education level was operationalized as having a four-year university degree or higher (or the country's equivalent), such that 0 = no degree and 1 = degree.

Income. Income was calculated by taking the median of a participant's income across their waves of data. This value was then transformed two ways within each dataset. First, it was standardized. Second, it was log-transformed to account for skewness of the raw income values. Specifically, the transformation $\log(x + 1)$ was used as some income values were zero.

Personality Traits. For each Big Five trait, the average value across all waves for a participant was calculated and standardized within each dataset. Traits were only included in the models in which the same trait was not the dependent variable (e.g., neuroticism was not treated as a covariate/moderator in the model with neuroticism as the dependent variable).

Life Events. The included life events were marriage, divorce, unemployment, university degree attainment, having a child, and experiencing a new health event (e.g., new health diagnosis). Broadly, life event variables were treated as dichotomous dummy-coded variables and coded 1 if an individual reported experiencing the event and coded 0 if not. Then, these life event variables were treated as both time-varying and time-invariant. For the time-varying, within-person variables, the coding of the variable was dependent upon if an individual first reported that life event at the given wave (0 if no, 1 if yes). These variables thus capture the *onset* of the event. For the time-invariant, between-person variables, if an individual reported a life event across any wave of data, they were coded as 1 for this variable and 0 if not. The time-varying and time-invariant variables were both included in every model for a given life event.

2.2.3 Outcome: Health Status

In order to conduct an initial test of the predictive utility of one's degree of within-person variability around their personality trajectory, a single outcome was chosen. Based on past work finding associations between levels and/or changes in all Big Five traits and self-report health status (Atherton et al., 2014; Hampson et al., 2006; Letzring et al., 2014; Magee et al., 2013; Takahashi et al., 2013; Wright & Jackson, 2022a, 2023a), and the inclusion of this variable in all datasets, this was the outcome of choice. The question for health status sometimes slightly varied across datasets and each item can be found in File S2. For each participant, their health status variable was either taken from the same wave as their final personality measure or from the nearest available distal wave (i.e., assessed after their final personality measure).

2.3 Transparency and Openness

Within this methods section, I report the final sample sizes and inclusion criteria, all measures and their psychometric properties, and follow the APA Style Journal Article Reporting Standards (JARS; Kazak, 2018). Data are accessible at all links specified in each dataset's "Participants" subsection. The codebook; R code for cleaning data, constructing variables, and running all analyses; and all supplementary materials are available on the OSF project page (https://osf.io/t7knz/). Data were analyzed using the package brms (Bürkner, 2017) in R (Version 4.2.1; R Core Team, 2021). The design and analyses were pre-registered (https://osf.io/x93e6).

2.4 Analytic Plan

The analytic plan consisted of four central steps: 1) testing measurement invariance across time for the Big Five; 2) fitting baseline MELSMs to examine if there were individual differences in longitudinal within-person residual variability, or sigma, for trajectories of the Big Five; 3) fitting additional MELSMs that included covariates to view if these factors were associated with heterogeneity in sigma; and 4) fitting a series of regression models to determine if person-level sigma values have predictive utility.

2.4.1 Measurement Invariance

First, to ensure meaningful variability in the constructs across time was being quantified, I conducted a series of measurement variance tests for each trait in each dataset. The R package semTools (Jorgensen et al., 2022) was used for all tests. The tested models were configural invariance, in which the same factor structure is imposed on all items across time; metric invariance, in which factor loadings are constrained to be equal on all items across time; scalar invariance, in which factor loadings and intercepts are constrained to be equal on all items across time; and residual invariance, in which factor loadings, intercepts, and residual variances are constrained to be equal on all items across time. Measurement invariance was evaluated using change in Confirmatory Factor Index (CFI) and change in Root Mean Square Error of Approximation (RMSEA) across models. Generally, criteria of -.01 change in CFI and .015 change in RMSEA were used to determine if each type of invariance was met across each successively more constrained model (Chen, 2007). However, given the sensitivity of these tests and of cut-off indices to factors such as sample size and model complexity, if a change in CFI or RMSEA greater than the specified cut-offs was found, the values of CFI and RMSEA for the tested model were considered as well. If the CFI was ≥ .95 and the RMSEA was ≤ .08, then the model was still considered to have acceptable fit (Hu & Bentler, 1999).

2.4.2 Modeling Individual Differences in Longitudinal Within-Person Variability

Second, after establishing measurement invariance, a series of baseline MELSMs were fit. These models quantify the degree to which individuals differ in their location (i.e., mean) *and* scale (i.e., variability) for each trait. MELSMs are an extension of standard MLMs such that they relax the assumption of homoscedasticity. This permits individuals to vary in their degree of within-person residual variability, which is quantified via the sigma parameter. Measurements across time (*i*, Level 1) were nested within individuals (*j*, Level 2). Separate models were fit for each trait within each dataset, resulting in a total of 25 models. An example equation can be demonstrated with the following:

Location:

$$Y_{ij} = b_{0j} + b_{1j} * time0_{ij} + e_{ij}$$
$$b_{0j} = \gamma_{00} + U_{0j}$$
$$b_{1j} = \gamma_{10} + U_{1j}$$

Scale:

$$\log(\sigma_{ij}) = \log\left(sd(e_{ij})\right) = \eta_{00} + w_{ij}$$

Where Y_{ij} is one of the trait variables; $time0_{ij}$ is centered at a participant's first measurement occasion (in units of one-year increments); and σ_{ij}, or $sd(e_{ij})$, is the person-level residual standard deviation for each trait, or the degree of within-person variability around one's trajectory. This latter quantity, sigma, is a combination of the sample-average variability of the person-level residual (η_{00}) and the person-specific deviation from this average variability (w_{ij}). The *i* and *j* subscripts indicate sigma can vary at the person (*j*) and occasion (*i*) level, thus permitting its value to change as a function of both time-invariant and time-varying covariates. Sigma is modeled with a log link function as it cannot be negative.

To evaluate if the MELSM was a meaningful improvement over a model that does not allow sigma to vary across individuals, standard MLMs were also fit. The specification of these MLMs was identical to that of the MELSMs, with the exception that sigma could not vary and thus random effects for this parameter were not included. As such, the MELSMs and MLMs were compared for each trait and dataset to determine if there were meaningful individual differences in sigma. The criteria used in model comparisons were Leave-One-Out Information Criteria (LOO-IC), obtained from LOO-Cross Validation (LOO-CV), and Watanabe-Akaike Information Criteria (WAIC). Generally, to conclude one model is meaningfully better, the difference between the expected log-predictive density (ELPD) estimates for the two models needs to be larger than four and more than double the magnitude of its standard error (Hollenbach & Montgomery, 2020; Johnson et al., 2022; Sivula et al., 2020), with some suggesting the ELPD difference needs to be as much as four times the magnitude of its standard error (Vehtari et al., 2017). The ELPD is an index of the model's estimated out-of-sample prediction accuracy, as quantified via LOO-IC or WAIC. Larger ELPDs are indicative of a better model and in a model comparison, the largest value is automatically scaled to 0. Thus, the model that has a negative value for the ELPD difference and a positive value for its standard error is the inferior model. Additionally, model weights are calculated from the information criteria and can be interpreted as the probability that a given model will be the best model, in terms of best out-of-sample prediction, ranging from 0 to 1 (Burnham & Anderson, 2002; Wagenmakers & Farrell, 2004). In the current study, to err on the conservative side, a model was considered the better-fitting model if the ELPD difference between it and the other model was larger than four and at least four times the magnitude of its standard error.

All models used weakly informative and regularized priors. For each dataset, the priors for intercepts were normal distributions with a mean equal to the nearest whole or half integer of the average Big Five trait value at the initial wave and a standard deviation of 1; the prior for the regression coefficient was a normal distribution with a mean of 0 and standard deviation of 1; the prior for the Level 2 standard deviation parameters (i.e., the random effects) was a half Cauchy distribution with a location of 0 and a scale of 2; and the prior for the sigma fixed effect was a log normal distribution with a location of 0 and a scale of 1. Maximum a posteriori (MAP) probability estimates, which are the mean values of the posterior, were obtained from each model's posterior distribution to serve as point estimates along with 95% credible intervals (CIs).

2.4.3 Extracting Individual-Level Parameter Values

After fitting the baseline MELSMs, the person-level intercept, slope, and sigma values were extracted from each model. For each of the 25 baseline MELSMs, the person-level sigma values were calculated using the fixed effect estimate and each individual's random effect across all samples in the posterior. The fixed effects are the mean values of the posterior distribution, and across all samples in the posterior distribution, each person's deviation from this fixed effect was added to it to obtain their own person-specific parameter value. In MELSMs, the use of a log link function to model sigma results in the model-provided fixed and random effects for sigma conveying information for *log-transformed* sigma. Thus, for each sample in the posterior, these person-specific values were then exponentiated – or backtransformed into their original units of the trait. Lastly, the median value across all samples was calculated for each person to give their final person-level parameter estimate (i.e., their individual-level sigma value). Then, each person's intercept and slope values were similarly obtained. With the exception of needing to exponentiate the model-provided parameter estimates, the person-level intercepts and slopes were calculated in an equivalent fashion to the person-level sigma values.

Oftentimes, these individual-level parameters provide useful information regarding the central tendency and dispersion of an effect within a sample, can be used to examine how different person-level parameters are associated with one another, and can be applied in further statistical tests. Although the estimated fixed and random effects in a model typically provide equivalent information as the person-level parameters, in MELSMs, the sample-level values for sigma, and thus their metrics of central tendency and dispersion, are quantified and calculated as logarithms. This results in the model-provided descriptive statistics for sigma (i.e., means, standard deviations, correlations) not necessarily directly corresponding with similar information instead obtained from the backtransformed (i.e., in original units), person-level sigma values.

This discrepancy is due to differences in how logarithmic and un/backtransformed values operate mathematically. Even simple calculations and descriptive statistics, such as the mean and standard deviation, will not always be equivalent, necessarily comparable, nor transformable from one metric to the other. Generally, arithmetic means for the individual-level sigma values are close approximations of the backtransformed, model-provided average sigma values (i.e., the fixed effects when exponentiated). However, the arithmetic medians and the (logs of the) geometric means for the person-level sigma values will be slightly better approximations of the backtransformed, sample-level sigma values. For the geometric mean of a set of values, its log is the average of those log values, thus it preserves the mathematical nuances of logarithms. For arithmetic medians, the median of backtransformed data will be the same as the median in the original (i.e., untransformed) data, and the median in the log scale will be the same as the mean in the log scale. That is, the model-provided, average log sigma values will be the same as their median log sigma values, and backtransforming these values will provide the same median but *not* the same mean as the original data.

Similarly, measures of dispersion cannot be transformed nor compared between logarithmic versus un/backtransformed data. Thus, the standard deviations for the person-level sigma values, regardless of the metric or method they were calculated by, will differ from the model-provided, sample-level estimates (i.e., the random effects for sigma). This is partly due to the use of the mean in the calculation of standard deviation, which changes over the course of transformation from original units to log units to backtransformation into original units. It is also due to the fact that what are additive relationships for logarithms are multiplicative relationships for un/backtransformed data. Thus, very different results are obtained when calculating what ± one standard deviation from the mean is when using logarithms versus un/backtransformed data.

For these reasons, I primarily focus on the person-level parameter values as these are in units of the trait and are the actual individual difference metrics of interest. The model-provided estimates for the fixed and random effects, along with their 95% CIs, do provide similar descriptive information, but it is for log-transformed sigma instead of the sigma that is in units of the traits. It is worth noting that because the person-level parameters in our study are themselves aggregated estimates (i.e., across all samples in the posterior), and thus inherently include variability, even metrics such as medians and geometric means of these person-level parameters will not be exactly equivalent to the backtransformed model-provided values.

2.4.4 Covariates of Individual Differences in Longitudinal Within-Person Variability

Next, I ran a series of MELSMs including covariates to examine if individual differences in longitudinal within-person variability, when present, could be explained by the inclusion of these variables. In these models, the person-specific sigma values are treated as outcomes and can be predicted by both Level 1 (i.e., within-person (*i*) level) and Level 2 (i.e., between-person (*j*) level) variables. Separate models were fit for each trait and dataset that had meaningful individual differences in sigma in its baseline MELSM (which ended up being all 25 models). Priors were the same as in the baseline MELSMs. Below is an example equation with a time-invariant covariate/moderator:

Location:

$$Y_{ij} = b_{0j} + b_{1j} * time0_{ij} + e_{ij}$$
$$b_{0j} = \gamma_{00} + \gamma_{01} * COVARIATE_j + U_{0j}$$
$$b_{1j} = \gamma_{10} + \gamma_{11} * MODERATOR_j + U_{1j}$$

Scale:

$$\log(\sigma_{ij}) = \log\left(sd(e_{ij})\right) = \eta_{00} + \eta_{01} * COVARIATE_j + w_{ij}$$

All models with only time-invariant variables (i.e., non-life-event models) were specified this way. In comparison, for the life event models, time-varying variables were in the Level 1 equation for the location parameters, as both a main effect and interaction with the time variable, whereas the time-invariant variables were in all Level 2 equations. Then, the equation for the scale parameter contained both the time-varying and time-invariant variables. Also, for the life event models, time was now centered around the time point immediately prior to someone reporting that life event. This meant that for individuals that did not report a given life event, time was centered around the average time point that individuals that did report the life event had. The intercepts in these models thus represent the average trait levels (location) and degree of within-person variability (scale) for individuals that did not experience a given life event.

2.4.5 Examining the Predictive Utility of Individual Differences in Sigma

Finally, to test the predictive utility of individual differences in sigma, a series of increasingly complex regression models in which each individual's sigma value predicted the outcome were fit. In all models, the outcome and predictor variables were standardized and the priors for all parameters were normal distributions with a mean of 0 and standard deviation of 1. For the first set of models, the most basic test is a simple regression in which only the person-level sigmas predict health status. An example equation can be demonstrated with the following:

$$Outcome_j = b_0 + b_1\sigma_j + e_j$$

Where σ_j is sigma, or an individual's degree of within-person variability around a given Big Five trait trajectory across time, and $Outcome_j$ is an individual's health status. Second, to serve as a more rigorous test, each person's intercept and slope values were also included as predictors in a multiple regression. An example equation can be seen below:

$$Outcome_j = b_0 + b_1\sigma_j + b_2Intercept_j + b_3Slope_j + e_j$$

Where $Intercept_j$ is an individual's initial value for a trait and $Slope_j$ is an individual's rate of change across time for a trait. Third, models with two-way interactions between each set of predictors were fit to examine how these individual differences may operate as a function of one another in the context of predicting health status, thus providing a comprehensive view of the effects. These models were not pre-registered. An example equation can be seen below:

$$Outcome_j = b_0 + b_1\sigma_j + b_2Intercept_j + b_3Slope_j +$$
$$b_4\sigma_j * Intercept_j + b_5\sigma_j * Slope_j + b_6Intercept_j * Slope_j + e_j$$

Lastly, for each of the three types of regression models, meta-analytic estimates were calculated. These analyses were not pre-registered but were conducted to obtain a concise summary of the results. Meta-analytic estimates, at least those obtained in this manner, were only calculated for the regression models as all other parameters of interest from the MELSMs (e.g., random effects for sigma, associations between sigma and the covariates) need to be interpreted in the context of other dataset-specific parameters (e.g., the sample-level sigma value). The parameters in the regression models are person-level and themselves standardized, though, which allows them to be compared across datasets. As such, for all regression models, parameters and standard errors for all predictors were extracted to obtain a weighted average of the effects – or the (estimated) true effect for each parameter. Multilevel models were then fit using these values, with estimates (*j*) nested in datasets (*k*). The below notation represents the true effect (μ) :

$$\hat{\theta}_k \sim N(\mu, \sigma_k^2 + \tau^2)$$

Where $\hat{\theta}_k$ is the "true" effect size in dataset *k*; N indicates the parameters were sampled from a normal distribution; μ represents the weighted, pooled true effect size of the *k* dataset-level effect size distributions; σ_k^2 is the variance of the effect size distribution for dataset *k*; and τ^2 is the variance of the distribution of the "true" effect sizes from the *k* datasets, such that it quantifies between-dataset heterogeneity. For all three types of regressions per Big Five trait, the true effect (μ_{00}) for each of its *j* parameters was estimated with the below equation:

$$Estimate_{jk} \mid SE_{jk} = \theta_{0k}$$
$$\theta_{0k} = \mu_{00} + U_{0k}$$

Where $Estimate_{jk}$ represents the observed effect size of parameter j from dataset *k*, weighted by its standard error (SE_{jk}); θ_{0k} represents the "true" effect size from dataset *k*, as its observed effect size has been corrected for error; μ_{00} represents the weighted, pooled true effect size; and U_{0k} represents the *k* dataset-specific deviations from the true pooled effect size. The meta-analytic estimates are in correlation units, just as the dataset-specific estimates are.

Chapter 3: Results

3.1 Measurement Invariance

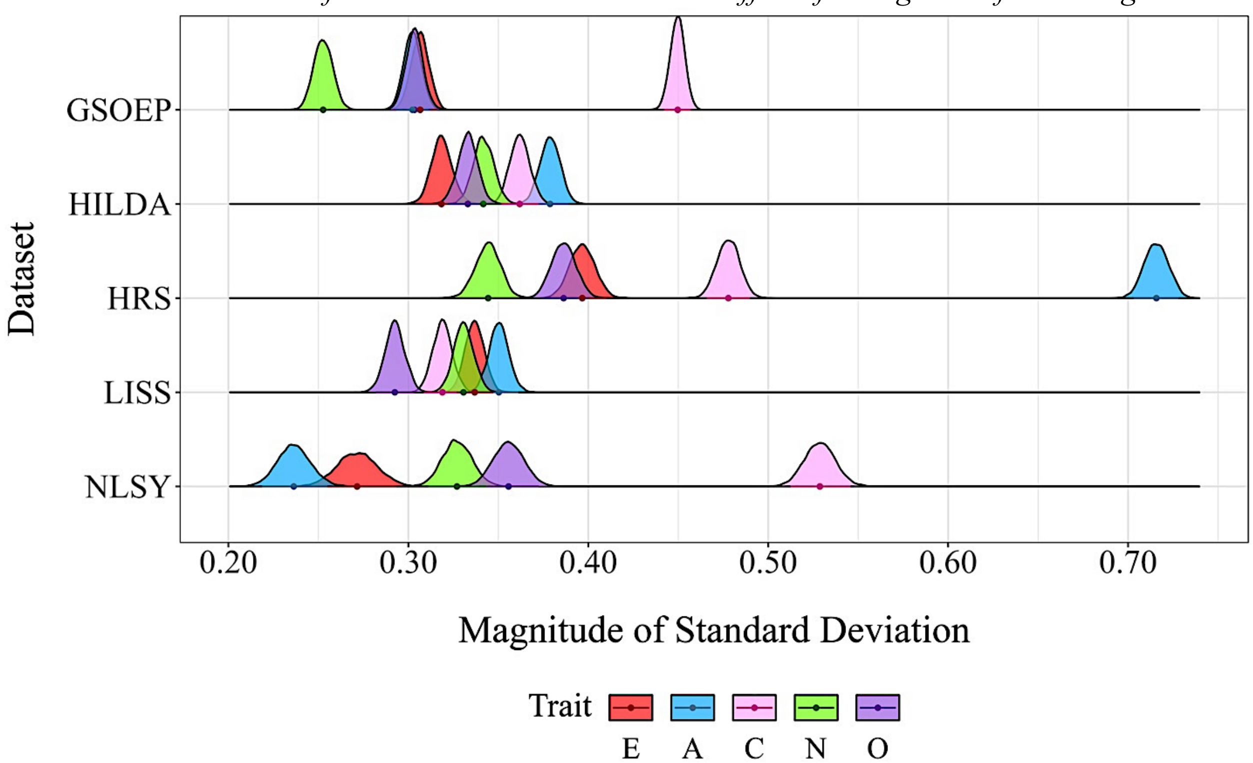
First, I ran a series of measurement invariance tests for each Big Five trait in each dataset. Acceptable model fit was found for all models. Full results from all measurement invariance tests are available in Table S2. For GSOEP, CFIs ranged from .967 to 1.000 and RMSEAs ranged from .004 to .025. Neuroticism was the only trait in which a CFI Δ exceeded -.01 (CFI Δ = -.019 for scalar). However, this model was still deemed acceptable as its CFI was .978 and its RMSEA was .014. For HILDA, CFIs ranged from .853 to .944 and RMSEAs ranged from .027 to .051. All measurement invariance tests were met. For HRS, CFIs ranged from .931 to .986 and RMSEAs ranged from .019 to .042. Conscientiousness was the only trait in which a CFI Δ exceeded -.01 (CFI Δ = -.016 for scalar). For this model, CFI was .966 and RMSEA was .025, thus it was still considered an acceptable model. For LISS, CFIs ranged from .909 to .956 and RMSEAs ranged from .020 to .032. All measurement invariance tests were met. For NLSY, CFIs ranged from .933 to 1.000 and RMSEAs ranged from .000 to .035. Conscientiousness was the only trait in which a RMSEA Δ exceeded .015 (RMSEA Δ = .017 for residual). The CFI of this model was .990 and its RMSEA was .018, thus it was likewise still considered acceptable.

3.2 Are There Individual Differences in Within-Person Variability Around Big Five Trajectories?

Next, I performed a series of baseline MELSMs to examine the degree to which there were meaningful individual differences in within-person residual variability for trajectories of the Big Five across time. Full results are available in Tables S3-S7. Across all models, there was meaningful variability in sigma (Table 2; Figure 6). This was true when judging variability in sigma according to if the 95% CI for its random effect (i.e., its standard deviation) included 0.00 and according to a model comparison in which the MELSM was compared to a MLM that did not allow sigma to vary. Parameter estimates from MLMs and results from model comparisons can be found in Tables S8-S12 (MLMs) and Tables S13-S17 (model comparisons). Of note, there was not a single instance in which a MELSM was not (overwhelmingly) the superior model compared to a standard MLM. Therefore, at the sample level, there is meaningful heterogeneity in within-person variability for Big Five trajectories – even when judged using multiple methods as well as when abiding by the more conservative guidelines for the criteria used in comparisons.

Figure 6

Posterior Distributions of the Model-Provided Random Effects for Log-Transformed Sigma



Note. The model-provided random effect for sigma is the degree of individual-level variability around the sigma fixed effect. Thus, the random effect is the variability, in standard deviation units, around the sample-level average value of *log-transformed* sigma in a given model.

Table 2

Model-Provided Estimates of the Fixed and Random Effects for Sigma

	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Dataset	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
GSOEP										
σ	1.03	[1.02, 1.04]	1.02	[1.01, 1.03]	0.97	[0.96, 0.97]	1.20	[1.19, 1.21]	1.13	[1.12, 0.14]
log(σ)	0.03	[0.02, 0.04]	0.02	[0.01, 0.03]	-0.03	[-0.04, -0.03]	0.19	[0.18, 0.19]	0.12	[0.12, 0.13]
$SD_{\log(\sigma)}$	0.31	[0.30, 0.32]	0.30	[0.29, 0.31]	0.45	[0.44, 0.46]	0.25	[0.24, 0.26]	0.30	[0.29, 0.31]
HILDA										
σ	0.77	[0.76, 0.78]	0.73	[0.72, 0.74]	0.82	[0.81, 0.83]	0.89	[0.88, 0.90]	0.89	[0.88, 0.90]
log(σ)	-0.26	[-0.27, -0.25]	-0.32	[-0.33, -0.30]	-0.20	[-0.21, -0.19]	-0.12	[-0.13, -0.11]	-0.11	[-0.12, -0.10]
$SD_{\log(\sigma)}$	0.32	[0.31, 0.33]	0.38	[0.37, 0.39]	0.36	[0.35, 0.37]	0.34	[0.33, 0.35]	0.33	[0.32, 0.34]
HRS										
σ	1.06	[1.04, 1.07]	0.79	[0.78, 0.80]	0.91	[0.90, 0.92]	1.14	[1.12, 1.15]	0.95	[0.93, 0.96]
log(σ)	0.05	[0.04, 0.07]	-0.23	[-0.24, -0.22]	-0.09	[-0.11, -0.08]	0.13	[0.11, 0.14]	-0.05	[-0.07, -0.04]
$SD_{\log(\sigma)}$	0.40	[0.38, 0.41]	0.72	[0.70, 0.73]	0.48	[0.47, 0.49]	0.34	[0.33, 0.36]	0.39	[0.37, 0.40]
LISS										
σ	0.61	[0.60, 0.62]	0.59	[0.58, 0.59]	0.59	[0.58, 0.60]	0.74	[0.74, 0.75]	0.55	[0.55, 0.56]
log(σ)	-0.50	[-0.51, -0.49]	-0.53	[-0.54, -0.52]	-0.53	[-0.54, -0.52]	-0.30	[-0.31, -0.29]	-0.59	[-0.60, -0.58]
$SD_{\log(\sigma)}$	0.34	[0.33, 0.35]	0.35	[0.34, 0.36]	0.32	[0.31, 0.33]	0.33	[0.32, 0.34]	0.29	[0.28, 0.30]
NLSY										
σ	1.58	[1.55, 1.62]	1.56	[1.53, 1.59]	1.30	[1.28, 1.33]	1.64	[1.61, 1.67]	1.51	[1.49, 1.54]
log(σ)	0.46	[0.44, 0.48]	0.45	[0.43, 0.46]	0.26	[0.25, 0.28]	0.49	[0.48, 0.51]	0.42	[0.40, 0.43]
$SD_{\log(\sigma)}$	0.27	[0.25, 0.29]	0.24	[0.22, 0.25]	0.53	[0.51, 0.55]	0.33	[0.31, 0.34]	0.36	[0.34, 0.37]

Note. Est = maximum a posteriori (MAP) estimate. CI = 95% credible interval. σ = backtransformed (i.e., exponentiated) sigma fixed effect. log(σ) = sigma fixed effect. $SD_{\log(\sigma)}$ = standard deviation around log(σ) (i.e., the random effect).

3.3 What is the Degree of Individual Differences in Longitudinal Within-Person Variability?

3.3.1 Person-Level Sigma Values

After fitting the baseline MELSMs, each of which indicated there was meaningful variability around the sample-level sigma value, I extracted the person-level intercept, slope, and sigma values. This allowed me to describe the extent of heterogeneity in the person-level sigmas, as well as the nature of their associations with the other person-level metrics, all in units of the trait. Table 3 contains descriptive information for these parameters across datasets. Table S18 contains further descriptive information for person-level sigmas, namely arithmetic, geometric, and logarithmic values for medians, means, standard deviations, and coefficients of variation.

Overall, the mean estimates of the person-level sigma values (Table 3) are comparable to their respective backtransformed (i.e., exponentiated) model-provided values (σ in Table 2). The model-provided, sample-level values are for log-transformed sigma, and logarithmic values differ in how they operate mathematically relative to un/backtransformed values. For central tendency metrics, means will be a close approximation, but medians and geometric means of the backtransformed, person-level sigma values will often be slightly more similar to the backtransformed, sample-level sigma value[1]. This is true even in light of the person-level parameters themselves being an aggregated estimate across all samples in the posterior. In this study, the means, medians, and geometric means were often a matter of some hundredths of a decimal apart (Table S18), though the medians often did provide the closest approximation of the backtransformed, sample-level sigma values.

[1] See the "Extracting Individual-Level Parameter Values" section in the Methods for more information.

Table 3

Descriptive Information for the Person-Level Intercept, Slope, and Sigma Values

	Extraversion			Agreeableness			Conscientiousness			Neuroticism			Openness		
Statistic	Int	Slope	Sigma	Int	Slope	Sigma	Int	Slope	Sigma	Int	Slope	Sigma	Int	Slope	Sigma
GSOEP															
M	6.60	-0.02	1.03	7.49	-0.01	1.03	8.19	-0.01	1.00	4.68	-0.03	1.21	6.12	-0.02	1.14
SD	1.38	0.02	0.16	1.01	0.01	0.20	0.86	0.01	0.38	1.44	0.03	0.11	1.44	0.02	0.18
CV (%)	20.93	96.42	15.32	13.48	89.72	19.02	10.50	43.57	37.73	30.79	83.04	9.29	23.63	86.48	15.55
M_{total}	–	-0.08	–	–	-0.07	–	–	-0.07	–	–	-0.15	–	–	-0.11	–
SD_{total}	–	0.19	–	–	0.15	–	–	0.08	–	–	0.28	–	–	0.22	–
HILDA															
M	5.91	-0.02	0.77	7.37	0.01	0.75	6.80	0.02	0.84	3.24	-0.02	0.90	5.38	-0.01	0.90
SD	1.36	0.02	0.12	1.06	0.01	0.19	1.26	0.02	0.19	1.21	0.02	0.20	1.38	0.02	0.17
CV (%)	23.00	92.24	15.25	14.40	116.35	25.17	18.53	82.09	23.31	37.45	88.54	22.14	25.58	106.63	18.35
M_{total}	–	-0.16	–	–	0.07	–	–	0.20	–	–	-0.17	–	–	-0.12	–
SD_{total}	–	0.27	–	–	0.21	–	–	0.27	–	–	0.27	–	–	0.27	–
HRS															
M	6.97	-0.03	1.08	8.66	-0.01	0.91	7.96	-0.02	0.96	3.51	-0.04	1.16	6.50	-0.03	0.97
SD	1.53	0.01	0.30	0.84	0.01	0.59	1.04	0.01	0.39	1.49	0.02	0.27	1.43	0.01	0.23
CV (%)	21.89	48.54	27.46	9.68	67.77	64.53	13.01	64.69	40.75	42.49	45.05	23.56	22.05	44.77	23.37
M_{total}	–	-0.18	–	–	-0.07	–	–	-0.12	–	–	-0.26	–	–	-0.21	–
SD_{total}	–	0.14	–	–	0.06	–	–	0.10	–	–	0.18	–	–	0.16	–
LISS															
M	5.71	-0.02	0.62	7.20	-0.01	0.60	6.69	0.00	0.60	4.04	-0.03	0.75	6.30	-0.01	0.56
SD	1.43	0.03	0.12	1.00	0.02	0.13	1.12	0.03	0.11	1.44	0.04	0.14	1.03	0.02	0.08
CV (%)	24.98	113.23	19.00	13.91	114.87	21.44	16.69	135.65	18.54	35.57	96.47	19.03	16.41	116.27	15.09
M_{total}	–	-0.10	–	–	-0.07	–	–	0.02	–	–	-0.16	–	–	-0.06	–
SD_{total}	–	0.33	–	–	0.25	–	–	0.29	–	–	0.40	–	–	0.21	–
NLSY															
M	6.23	-0.05	1.60	6.67	0.02	1.57	7.75	0.05	1.38	3.28	-0.04	1.66	7.53	-0.02	1.54
SD	1.29	0.04	0.25	0.84	0.03	0.26	1.03	0.02	0.61	1.07	0.02	0.43	0.81	0.01	0.41
CV (%)	20.68	66.98	15.90	12.60	87.86	16.83	13.35	46.18	43.82	32.55	56.99	25.65	10.80	42.20	26.94
M_{total}	–	-0.37	–	–	0.17	–	–	0.34	–	–	-0.27	–	–	-0.17	–
SD_{total}	–	0.32	–	–	0.23	–	–	0.19	–	–	0.19	–	–	0.10	–

Note. M = mean. *SD* = standard deviation. CV = coefficient of variation, expressed as a percentage. Int = intercept. Values are in the original or backtransformed units of the trait. Descriptive information was calculated with arithmetic mathematical operations using these values.

I next looked at the coefficient of variation (CV). The CV is a standardized measure of dispersion that provides information about the degree of variability in a variable relative to its average value. The CV is a standardized, unitless index and relative measure of dispersion. A relative measure (e.g., CV) differs from an absolute measure (e.g., standard deviation) of dispersion as the latter instead retains its variable's units, making cross-variable comparisons difficult. To calculate a CV, a variable's standard deviation is divided by its mean. This value can further be multiplied by 100 so the CV can be interpreted as the percentage that a variable's magnitude of spread is relative to the magnitude of its mean. Lower CVs signify that a variable has little variability compared to its mean value, with 0% being the minimum possible value and indicating there is no meaningful heterogeneity whatsoever. As such, the variable's mean would provide all the information needed to describe it. In comparison, higher CVs imply a variable has greater variability. For instance, a CV of 25% indicates the standard deviation of a variable is a quarter of its mean and a CV above 100% would indicate a variable's magnitude of variability is larger than the magnitude of its mean. Simply put, as a CV gets further in value from 0%, there exists a greater amount of meaningful heterogeneity in the variable. Accordingly, ignoring this variability and instead only using its average value, such as is done when not permitting there to be individual differences in a model parameter, results in a loss of information.

When examining CVs for the sigma values, there is clearly meaningful variability (Table 3). The CVs ranged from 9.29% in GSOEP for neuroticism to 64.53% in HRS for agreeableness, with a grand average CV of 24.12%. Thus, across five datasets that vary in methodology, design, average age, and country of origin, the sigma values, on average, vary in magnitude to the degree of just under 25% of their mean value. This degree of variability would be equivalent to a level of a trait – a metric of personality with well-established individual differences that one would be criticized for not accounting for – having a mean and standard deviation of, for example, 6 and 1.5 or 5 and 1.25, respectively (see Table 3 for an idea of the plausibility of these values).

3.3.2 Comparing Heterogeneity in Person-Level Intercept, Slope, and Sigma Values

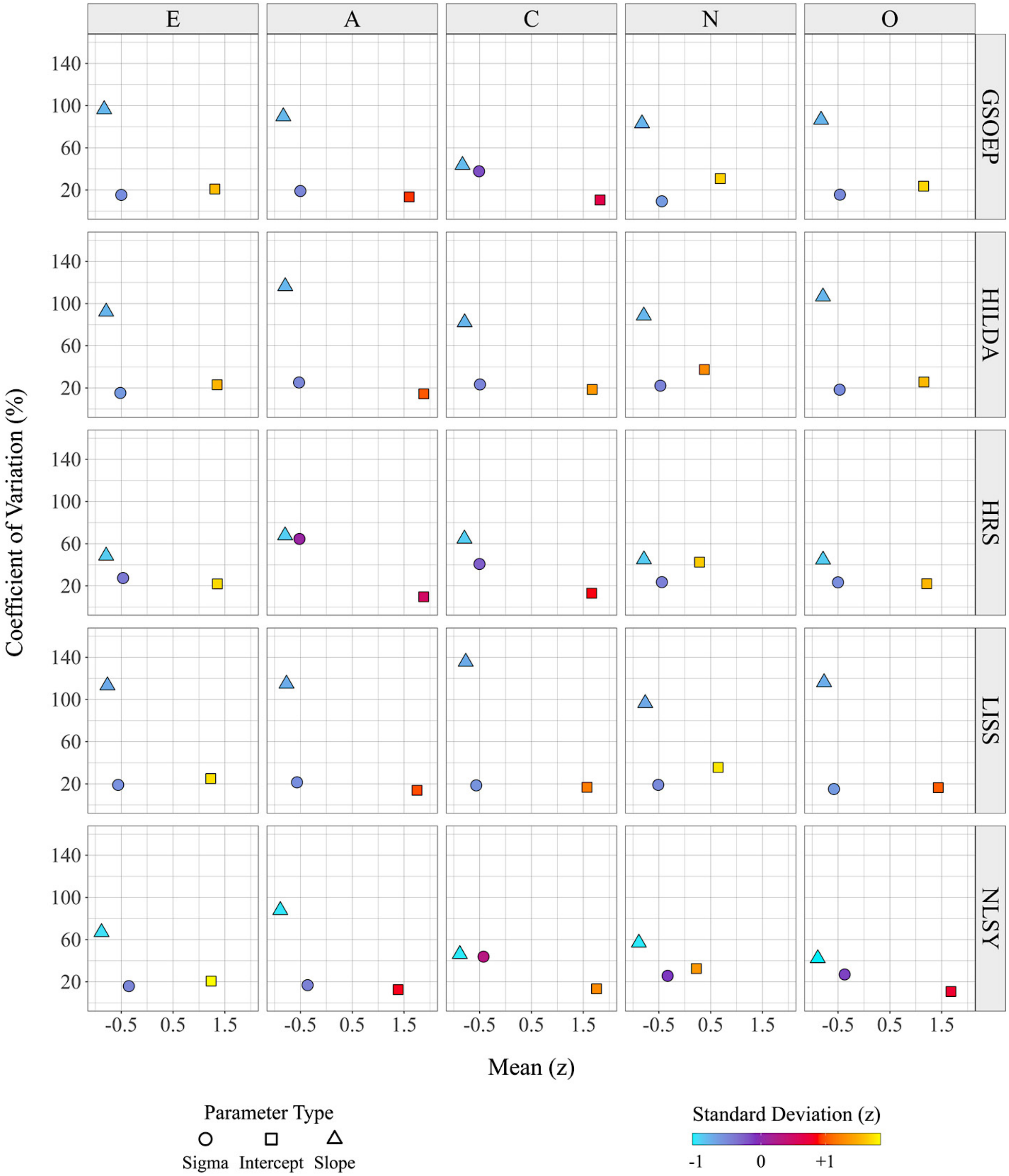
The presence of individual differences in longitudinal within-person variability around Big Five trajectories is a noteworthy finding in and of itself. However, and as foreshadowed two sentences ago, comparing the magnitude of the individual differences in sigma to that of other more frequently examined individual differences in personality development can help contextualize their degree of heterogeneity. Specifically, individual differences in sigma values can be compared to individual differences in intercepts (i.e., initial trait levels) and slopes (i.e., changes in traits). If there are comparable levels of variability in sigma as there are intercepts and slopes – two individual differences that are focal parameters of interest in personality development research – then it suggests ignoring this heterogeneity in sigma could be akin to assuming all people have the same initial trait levels or change in the exact same way across time. Notably, both are assumptions most researchers (rightfully) would not make. Thus, using the degree of variability in intercepts and slopes as a benchmark for comparison with variability in sigma values serves to ascribe meaning to their heterogeneity and highlight their potential theoretical and empirical utility. In the current study, these comparisons were done in two ways.

Standardized Measure of Dispersion. First, the CVs for all parameters were compared (Table 3; Figure 7). Since the parameters are not all quantified in the same units, raw descriptive metrics, such as their standard deviations, cannot be directly compared. However, CVs can be used to compare the degree of heterogeneity in all three parameters, regardless of their units, as it is a standardized, relative measure of dispersion. It should be noted that for slopes, the CV was calculated using their original standard deviation, but the mean slope was calculated from the absolute values of the slopes. Standard deviation is unaffected when there are both positive and negative numbers, so its original value is accurate, but the *magnitude* of the mean will be underestimated. Thus, calculating the CV for the slopes in this manner (which is sometimes referred to as the relative standard deviation) permitted the most accurate estimate of this metric.

Table 3 contains the CVs for all parameters across all traits and datasets, while Figure 7 further provides a graphical display of these values. Generally, the largest CVs are found for parameters with the smallest means (e.g., the slopes in Figure 7, which are depicted as triangles). A parameter with a larger mean can offset the magnitude of its mean if it also has a larger standard deviation, though, as this results in a higher CV (as seen for the extraversion square compared to the agreeableness square in the NLSY row of Figure 7). Then, in terms of the actual CV values, for intercepts, the average CV was 19.87% in GSOEP, 23.79% in HILDA, 21.82% in HRS, 21.51% in LISS, and 18.00% in NLSY – with a simple average across all datasets of 21.00%. The weighted average, which takes the sample size of each dataset into account, was 21.06%. For slopes, the average CV was 79.85% in GSOEP, 97.17% in HILDA, 54.16% in HRS, 115.30% in LISS, and 60.04% in NLSY – with a grand simple average of 81.30% and weighted average of 80.84%. For sigma values, the average CV was 19.38% in GSOEP, 20.84% in HILDA, 35.94% in HRS, 18.62% in LISS, and 25.83% in NLSY – with a grand simple average of 24.12% and weighted average of 23.13%. From these numbers, it is apparent that person-level slopes have the largest ratios of their standard deviations to their means. However, the intercepts and sigmas have extremely similar standardized degrees of dispersion, and, on average, the sigma values have a larger CV than the intercepts.

Figure 7

Distribution of the Coefficients of Variation for Person-Level Intercept, Slope, and Sigma Values



Note. To ensure plot elements for the different person-level parameters were on similar scales, their means and standard deviations were standardized (z) and used **only** for plotting, not for any calculations. Values were standardized across all traits and parameters within each dataset. E = extraversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

Unstandardized Measure of Dispersion. Table 3 also contains means and standard deviations for the *total* slope a person has across their waves of data. Variables having different units prevents meaningful comparisons of their unstandardized descriptive metrics, such as standard deviations. Intercepts and sigmas are in units of the traits, but slopes are in units of change in a trait per change in one year. To compare the standard deviations of slopes with those of intercepts and sigmas, the slopes must be transformed to where they quantify the *total* slope a person has across their waves of data – effectively the total amount of change they had in a trait. Each person provided data for a certain number of waves, and the timespan over which they provided data can vary in duration relative to other people. Their slope reflects the average amount of linear change they had across their waves of data, notwithstanding some shrinkage towards the average slope value, scaled in one-year increments. If this slope is multiplied by the total number of years they provided data, then an estimate of their *total* slope or change in a trait is provided. Importantly, this sets the slope on the same time scale that sigmas were quantified over. This renders the units of time irrelevant and permits direct comparisons of their descriptive information with intercepts and sigmas. For those individuals with only one wave of data, their slope estimate was multiplied by the average number of years people in a study had for that trait.

For the intercepts, the average standard deviations ranged from 1.01 in NLSY to 1.26 in HRS – with an overall simple average of 1.19 and weighted average of 1.22. For the *total* slopes, the average standard deviations ranged from 0.13 in HRS to 0.29 in LISS – with a grand simple average of 0.21 and weighted average of 0.20. For the sigma values, the average standard deviations ranged from 0.12 in LISS to 0.39 in NLSY – with a grand simple average of 0.25 and weighted average of 0.23. Clearly, the greatest absolute variability exists for initial trait levels. However, the magnitude of the standard deviations for total changes in a trait and sigma values is extremely similar. In almost half of the models (44%), sigma values had larger variability than total slopes. Furthermore, on average, the standard deviations for sigma were larger than those for total changes in traits. This indicates that individual differences in sigma exhibit a similar magnitude of absolute variability that total changes in personality across multiple years do.

3.3.3 Summary

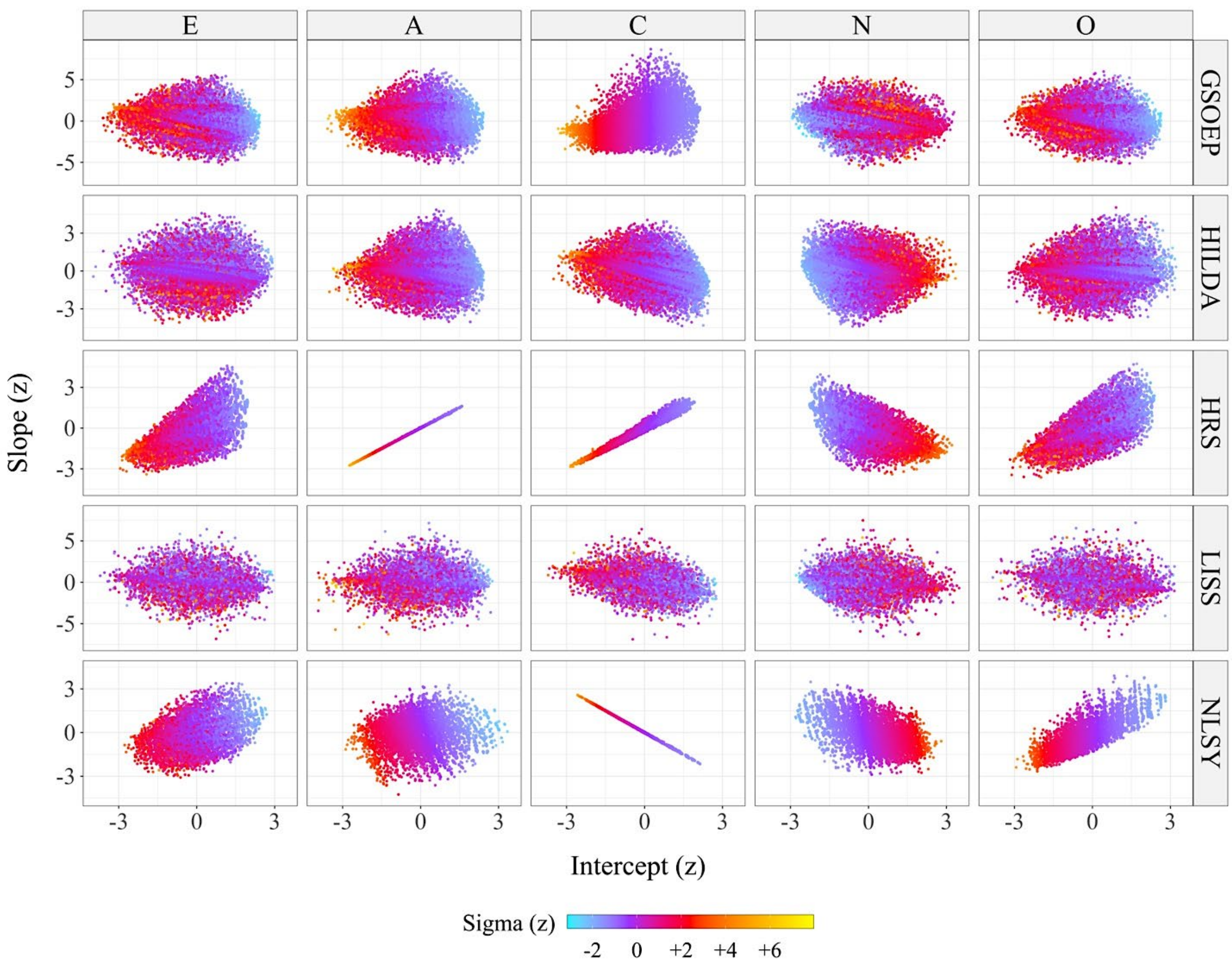
In sum, conclusions regarding the presence and degree of individual differences in longitudinal within-person variability around Big Five trajectories, regardless of if derived from the sample-level estimates or the person-specific sigma values, do not change. A standardized metric of dispersion indicated that sigma values, on average, have a larger degree of relative variability than intercepts – the latter of which is variability that is, by default, quantified in standard MLMs and considered to be a valuable source of between-person differences for personality development research. Similarly, when converted into the same units, the average magnitude of the standard deviations for sigma values was larger than that of total changes in personality traits across multiple years. Overall, these results highlight that there is considerable heterogeneity in the degree of within-person variability around individuals' Big Five trajectories, a finding that replicates across five datasets and for measures of both relative and absolute dispersion. Lastly, these findings importantly show that the magnitude of this variability is comparable to that of frequently examined individual differences in personality development.

3.4 Are Individual Differences in Sigma Associated with Other Personality Individual Differences?

Next, I examined how individual differences in longitudinal within-person variability for a trait were associated with other personality individual differences. Specifically, I examined how sigma values for a Big Five trait were associated with 1) with levels and changes in the same trait (Table 4; Figures 8-9) and 2) sigma values for all other traits (Table 5).

Figure 8

Scatterplots of the Associations Between Person-Level Intercept, Slope, and Sigma Values



Note. Standardized (z) values are plotted for each person-level parameter. Standardization was done within each dataset, trait, and parameter combination. E = extraversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness.

Table 4
Correlations Between Person-Level Intercept, Slope, and Sigma Values

	Level-Change					Level-Sigma					Change-Sigma				
Dataset	E	A	C	N	O	E	A	C	N	O	E	A	C	N	O
GSOEP	-.48	-.42	.45	-.49	-.48	-.74	-.92	-.98	.49	-.78	.30	.30	-.44	-.13	.26
HILDA	-.14	-.10	-.42	-.22	-.02	-.09	-.77	-.78	.83	-.56	-.24	-.03	.29	-.06	-.06
HRS	.65	1.00	.99	-.73	.67	-.89	-.95	-.97	.92	-.75	-.64	-.95	-.95	-.72	-.62
LISS	-.20	-.10	-.48	-.21	-.25	-.07	-.45	-.50	.44	.08	-.19	-.19	.28	-.06	-.08
NLSY	.36	.00	-1.00	-.51	.78	-.85	-.96	-.97	.98	-.98	-.44	-.24	.97	-.41	-.77

Note. Correlations between the person-level parameters of initial levels of traits, changes in traits (i.e., slopes), and within-person residual variability, or sigma, are presented for each Big Five trait in each dataset. E = extraversion. A = agreeableness. C = conscientiousness. N = neuroticism. O = openness. All correlations above $r = |\ .02\ |$ are significant at $p < .001$. All correlations above $r = |\ .01\ |$ are significant at $p < .01$.

Table 5
Correlations Between Person-Level Sigma Values Across All Big Five Traits

	GSOEP				HILDA				HRS				LISS				NLSY			
Trait	E	A	C	N	E	A	C	N	E	A	C	N	E	A	C	N	E	A	C	N
A	.09	–			.18	–			.52	–			.34	–			.10	–		
C	.16	.28	–		.19	.37	–		.43	.50	–		.30	.37	–		.14	.19	–	
N	.16	.09	.03	–	.19	.41	.39	–	.26	.17	.28	–	.34	.26	.30	–	.14	.25	.32	–
O	.32	.13	.12	.14	.19	.20	.15	.08	.50	.42	.48	.23	.34	.32	.32	.27	.27	.21	.24	.20

Note. Correlations between the person-level sigma values across all Big Five traits are presented for each dataset. E = extraversion. A = agreeableness. C = conscientiousness. N = neuroticism. O = openness. All correlations are significant at $p < .001$.

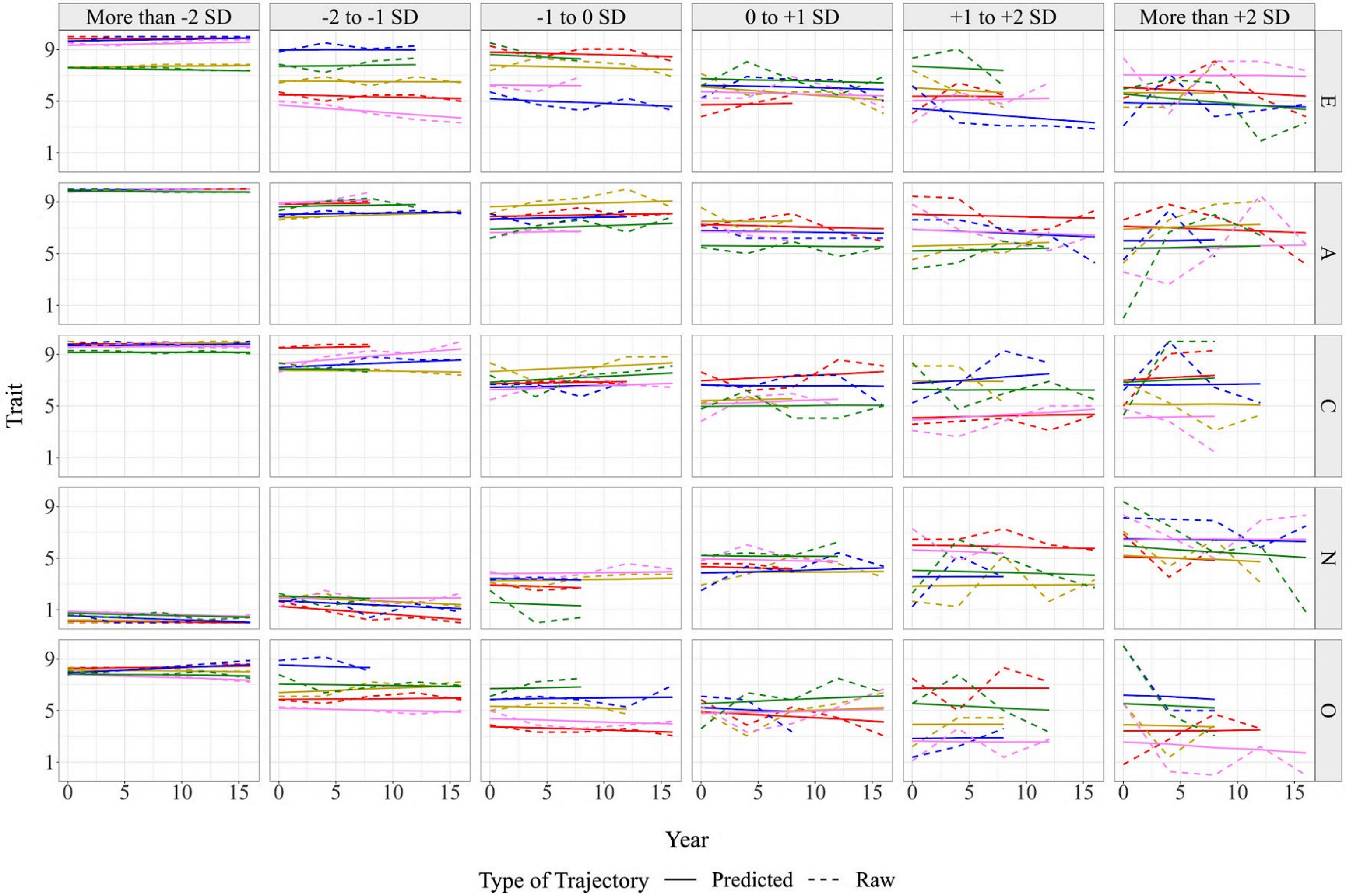
3.4.1 Levels of Traits

When examining the correlation between intercept and sigma values for the same Big Five trait (Table 4; Figure 8), the general pattern was that longitudinal within-person variability for a trait was negatively correlated with its initial levels for all traits except neuroticism. For neuroticism, the intercept and sigma values were instead positively correlated. The simple average correlation across datasets was $r = -.53$ for extraversion; $r = -.81$ for agreeableness; $r = -.85$ for conscientiousness; $r = .73$ for neuroticism; and $r = -.60$ for openness. The weighted average correlation was $r = -.56$ for extraversion; $r = -.84$ for agreeableness; $r = -.88$ for conscientiousness; $r = .66$ for neuroticism; and $r = -.63$ for openness. These correlations indicate that if someone has a higher initial level of extraversion, agreeableness, conscientiousness, or openness, they will, on average, have less within-person variability around their trajectory for that trait relative to people with lower initial levels. Then, the opposite is true for neuroticism.

Furthermore, these associations can be demonstrated graphically. To better understand the patterns of change in the Big Five traits – or the nature of the trajectories – for individuals with different degrees of longitudinal within-person variability, I plotted a random sample of individual-level trajectories in each dataset using the model-implied, predicted values as well as the raw data (Figure 9; Figures S1-S5). This allowed for a closer examination of how individual differences in sigma manifested in personality trajectories across time. For instance, in Figure 9, as the magnitude of sigma for extraversion, agreeableness, conscientiousness, and openness gets larger, these individuals tend to also have lower initial trait levels. Conversely, the opposite is true for neuroticism (Figure 9). This corresponds with the correlations found between intercepts and sigma values in these traits (Table 4).

Figure 9

Individual-Level Trajectories of Predicted and Raw Big Five Scores Plotted as a Function of Sigma Values



Note. The graphs for each trait are split into six panels, faceted by the magnitude of person-level sigma values in standard deviation units. There are five randomly sampled participants in each panel. A participant's predicted and raw trajectories are plotted in the same color. E = extraversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness. Data are from HILDA.

3.4.2 Changes in Traits

Next, correlations between slopes and sigmas were largely consistent for all traits except conscientiousness (Table 4; Figure 8). Generally, the value of a slope was opposite of the amount of sigma it was related to. Larger systematic increases (or smaller decreases) were associated with less variability around a trait's trajectory. Conversely, the more someone decreased (or increased less) in a trait, they often also had greater variability around that trajectory. However, for conscientiousness, the opposite pattern was mostly found: larger increases were associated with larger sigma values and larger decreases were associated with smaller sigma values.

Across all datasets, the simple average correlation was r = -.24 for extraversion; r = -.22 for agreeableness; r = .05 for conscientiousness; r = -.28 for neuroticism; and r = -.25 for openness. Then, the weighted average correlation was r = -.09 for extraversion; r = -.09 for agreeableness; r = -.21 for conscientiousness; r = -.24 for neuroticism; and r = -.07 for openness. As for a graphical example of these associations, in Figure 9, as sigma values for extraversion and openness get larger, the predicted trajectories also appear to have steeper decreases.

Additionally, it is worth noting that even for individuals with an average or below-average sigma value for a trait, it is rarely the case that they change in a constant, linear fashion (Figure 9; Figures S1-S5). Someone having little variability around their trajectory does not imply they change in a way that reflects their linear line of best fit; rather, the spread of their data points around this line is simply more condensed relative to those with greater variability. For individuals with smaller sigmas, their line of best fit appears to at least mostly be an acceptable approximation of how they are changing in a trait across time – albeit still an oversimplification. In comparison, for individuals with greater-than-average sigmas, the spread of data points around their lines of best fit is substantially more widespread and these lines themselves are typically a poor depiction of the true nature of their development. Moreover, this discrepancy between one's linear line of best fit and their true pattern of change across time is not only overlooked when using typical models, it cannot even be quantified. However, this is precisely what is done in MELSMs – further highlighting their potential utility in personality development research.

3.4.3 Sigma Values for Other Big Five Traits

Lastly, a clear pattern emerged when examining if sigma values in one trait were associated with sigma values in other traits (Table 5). Across all datasets and traits, individuals that were more variable around their trajectory for one trait were, on average, more variable for all other traits. The simple average correlation for sigma values in a trait with sigma values in all other traits was $r = .26$ for extraversion; $r = .27$ for agreeableness; $r = .28$ for conscientiousness; $r = .23$ for neuroticism; and $r = .26$ for openness. The weighted average correlation was $r = .25$ for extraversion; $r = .25$ for agreeableness; $r = .25$ for conscientiousness; $r = .19$ for neuroticism; and $r = .24$ for openness. Overall, these results suggest having greater longitudinal within-person variability might not only be a trait-specific individual difference, but further that it could be one that is more so a property of the person that, accordingly, manifests for each Big Five trait.

3.5 Which Variables Are Associated with Within-Person Variability Around Big Five Trajectories?

Next, I examined if there were variables associated with heterogeneity in sigma. This was done through a series of MELSMs for each trait, covariate, and dataset. Overall, variables of central empirical and theoretical importance in personality development research were meaningfully associated with heterogeneity in sigma values for Big Five traits. These covariate-sigma associations notably emerged above and beyond any effects rather due to the variable's associations with intercepts or slopes. Thus, these effects emerged even after a conservative test.

To interpret the effects, the backtransformed value of the average sigma in a model (for when a covariate = 0; i.e., the sigma fixed effect) can be compared to the backtransformed value of sigma for a one-unit change in a covariate. The backtransformed, model-average sigma is obtained via exponentiating the sigma fixed effect. Then, the backtransformed sigma for a one-unit change in a covariate is calculated by 1) adding (or subtracting) the sigma fixed effect and the fixed effect for the covariate's association with sigma and 2) exponentiating this value.

For an effect size, the average percent change in backtransformed sigma for a one-unit increase in a covariate is provided for each covariate and trait combination (Table 6). Although this effect size includes the word "change," this does not imply nor is it meant to be interpreted as the average percent that individuals' own sigma values changed as a function of a covariate. For all time-invariant covariates (i.e., all covariates except the time-varying life event variables), it simply quantifies the percentage that the average value of backtransformed sigma for when a covariate = 1 differs from the backtransformed, model-average sigma (i.e., when a covariate = 0), relative to the magnitude of this model-average sigma. For time-varying life event covariates, though, it can be interpreted as the average percentage that people's own sigma values changed when they experienced a life event, relative to their pre- and post-event average sigma values.

All covariate-sigma associations are available in Table 7 (gender, age, cognitive ability, education level, income), Table 8 (personality traits), and Tables 9-10 (life events). All model-provided estimates can be found in Tables S19-S36. The percent change estimates for individual models are available in Table S37.

Table 6

Average Percent Change in the Backtransformed, Model-Provided Sigma Value for a One-Unit Increase in a Covariate per Big Five Trait

Covariate	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Gender	**-1.36**	**-11.51**	**-4.80**	**4.11**	-0.58
Age[1]	**-7.36**	**-5.89**	**-8.85**	-6.11	**-2.96**
Age[2]	**10.10**	**9.12**	**13.70**	**6.47**	**7.53**
Cognitive Ability	**-11.19**	**-13.15**	**-12.72**	**-12.10**	**-16.49**
Education Level	**-12.28**	**-13.65**	**-13.12**	**-13.43**	**-16.03**
Income (log)	**-3.53**	**-1.78**	**-5.04**	**-3.33**	**-4.28**
Extraversion	–	**-8.82**	**-6.11**	**-1.52**	**-3.62**
Agreeableness	**-2.80**	–	**-10.40**	**-2.44**	**-3.06**
Conscientiousness	**-4.73**	**-10.23**	–	**-3.56**	**-4.14**
Neuroticism	**5.59**	**9.33**	**12.43**	–	**4.14**
Openness	**-5.40**	**-8.93**	**-8.00**	**-0.94**	–
Marriage (B)	**-11.23**	**-9.32**	**-14.97**	-6.06	-3.80
Marriage (W)	2.06	0.72	1.03	3.48	2.24
Divorce (B)	-0.19	**-1.55**	**-4.09**	**2.29**	0.44
Divorce (W)	**11.47**	3.05	3.32	8.01	6.77
Unemployment (B)	**10.54**	**7.95**	**14.43**	**10.77**	**11.65**
Unemployment (W)	**6.90**	3.95	**5.58**	**7.27**	4.55
Degree Attainment (B)	**-12.11**	**-14.48**	**-10.53**	**-13.32**	**-14.97**
Degree Attainment (W)	**7.61**	**6.67**	6.30	9.11	**5.80**
Child (B)	**-6.12**	**-2.65**	**-10.94**	**-3.44**	**0.05**
Child (W)	5.96	2.84	3.89	5.97	3.06
Health Event (B)	**0.76**	**-0.28**	1.35	2.04	**0.71**
Health Event (W)	1.43	1.65	2.26	3.48	2.08

Note. A positive value means a one-unit increase in a covariate was associated with a *larger* sigma and a negative value means it was associated with a *smaller* sigma. The values for the age covariates are from models including linear and quadratic age. The between-person, time invariant life event variables are marked with a "(B)" and within-person, time-varying variables are marked with a "(W)". A "–" in a cell indicates those models were not run for that covariate and trait combination. Bolded values are those in which the effect of the covariate was meaningful in at least half of the datasets.

3.5.1 Gender

For gender (Table 7; Table S19), females had less longitudinal within-person variability around their trajectories than males did for extraversion, agreeableness, and conscientiousness. The opposite effect was found for neuroticism, such that females had larger sigma values. Effects were generally small in magnitude, with female-average sigma values differing from male-average sigma values an average of -1.36% for extraversion, -11.51% for agreeableness, -4.80% for conscientiousness, 4.11% for neuroticism, and -0.58% for openness.

3.5.2 Age

Generally, for all datasets except HRS, the linear effect of age was negative in direction for all traits (Table 7). This suggests that the older people are, the less within-person variability they have around their trajectories. Although, for these same datasets, the quadratic effect was positive, which indicates the linear effect attenuates at older ages. For instance, individuals that are a few years apart in age (45 versus 42) but near the average age (40) may differ in sigma values by -0.06, whereas people much older than average (75 vs 72) may only differ by -0.01. Further, in models where the quadratic effect is larger in magnitude than the linear effect, older individuals will eventually have larger sigmas than people only slightly younger than them, thus nullifying the negative linear effect completely. Then, in HRS, the linear and quadratic effects were both positive for all traits. This indicates that the older people are, the larger their sigma values are compared to younger individuals. Further, the magnitude of this difference in sigmas is compounding in nature, such that it is largest in older ages and smallest near the average age.

For linear age effects, the average percent change in sigma for people one standard deviation larger than the average age relative to those of average age was smallest in magnitude for GSOEP (-0.74%) and largest for NLSY (-16.98%). For quadratic age effects, the average percent change was smallest in HRS (4.50%) and largest in NLSY (22.42%). Trait-specific effects were, on average, largest for conscientiousness (-8.85% for linear age and 13.70% for quadratic age) and smallest for openness (-2.96% for linear age and 7.53% for quadratic age).

Overall, results across datasets are consistent with one another due to HRS having an average age that is older than other datasets. Younger people tend to have greater within-person variability and people that are near or slightly older than middle age tend to have less variability. Older individuals will either taper off in having less variability or often again instead start to have larger sigma values relative to their slightly younger companions. These results suggest sigma as a function of age may resemble a U-shaped curve. See Table S20 for results from models that only included linear age and Table S21 for the models with linear and quadratic age.

3.5.3 Cognitive Ability

Across all traits and datasets, higher cognitive ability was associated with less within-person variability around trajectories (Table 7; Table S22). The magnitude of effects was comparable for all traits, with openness having the largest average percent change in sigma (-16.49%). Effects were particularly large in HRS, such that having one standard deviation higher-than-average cognitive ability scores was associated with a sigma an average of 20.61% smaller than the model-average sigma (i.e., sigma for average cognitive ability).

3.5.4 Education Level

Having a university degree was associated with less longitudinal within-person variability for all traits across datasets (Table 7; Table S23). Effects were quite large for HRS and NLSY, such that having a university degree was associated with a sigma an average of 17.11% (HRS) and 19.24% (NLSY) smaller than the model average. Across traits, effects were similar in magnitude and openness had the largest average percent change in sigma (-16.03%).

3.5.5 Income

For log-transformed income, larger values were nearly always associated with a smaller sigma value for all Big Five traits (Table 7). Since income and sigma are both log transformed, the model-provided coefficient for income's association with sigma can be interpreted as the percent increase in log sigma for every 1% increase in log income. To calculate the percent increase in log sigma for any desired percent increase in log income: raise the desired percent increase (x, in decimal form) to the power of the coefficient (c); subtract one; and multiply by 100 (i.e., $(x^c - 1) * 100$). As for the average percent changes, a one-unit increase in log income corresponds to an income that is larger by a factor of e^1 (approximately 2.72). For this difference in income values, effects were generally small in magnitude across traits, with an average percent change in sigma of -3.53% for extraversion, -1.78% for agreeableness, -5.04% for conscientiousness, -3.33% for neuroticism, and -4.28% for openness.

See Table S24 for results from models with the standardized income variable and Table S25 for all estimates from the models with log-transformed income.

Table 7

Associations of Gender, Age, Cognitive Ability, Education Level, and Income as Covariates with Within-Person Residual Variability for Trajectories of the Big Five Traits

	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Model	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
Gender										
GSOEP	**-0.02**	**[-0.04, -0.01]**	**-0.07**	**[-0.09, -0.06]**	**-0.08**	**[-0.10, -0.07]**	**0.05**	**[0.03, 0.06]**	0.01	[-0.00, 0.03]
HILDA	0.02	[0.00, 0.04]	**-0.10**	**[-0.12, -0.08]**	-0.01	[-0.03, 0.01]	-0.00	[-0.02, 0.01]	0.00	[-0.01, 0.02]
HRS	**-0.03**	**[-0.06, -0.01]**	**-0.38**	**[-0.41, -0.36]**	**-0.10**	**[-0.12, -0.08]**	**0.04**	**[0.02, 0.06]**	-0.01	[-0.03, 0.01]
LISS	0.01	[-0.01, 0.03]	-0.02	[-0.04, -0.00]	-0.00	[-0.02, 0.02]	**0.04**	**[0.02, 0.06]**	0.01	[-0.01, 0.03]
NLSY	**-0.05**	**[-0.08, -0.02]**	**-0.08**	**[-0.11, -0.06]**	**-0.06**	**[-0.10, -0.03]**	**0.07**	**[0.04, 0.10]**	**-0.04**	**[-0.07, -0.01]**
Age[1] & Age[2]										
GSOEP	**-0.02**	**[-0.03, -0.01]**	0.01	[-0.00, 0.01]	**-0.06**	**[-0.07, -0.05]**	-0.01	[-0.02, -0.00]	**0.04**	**[0.03, 0.04]**
(Age^2)	**0.05**	**[0.05, 0.06]**	0.01	[0.00, 0.02]	**0.11**	**[0.10, 0.12]**	**0.04**	**[0.03, 0.04]**	**0.04**	**[0.03, 0.05]**
HILDA	**-0.08**	**[-0.09, -0.07]**	**-0.09**	**[-0.10, -0.08]**	**-0.09**	**[-0.10, -0.08]**	**-0.10**	**[-0.11, -0.09]**	**-0.08**	**[-0.09, -0.07]**
(Age^2)	**0.10**	**[0.09, 0.12]**	**0.11**	**[0.10, 0.12]**	**0.11**	**[0.10, 0.12]**	**0.07**	**[0.06, 0.08]**	**0.11**	**[0.10, 0.12]**
HRS	**0.03**	**[0.02, 0.04]**	**0.04**	**[0.03, 0.05]**	**0.09**	**[0.08, 0.10]**	-0.01	[-0.02, -0.00]	**0.05**	**[0.03, 0.06]**
(Age^2)	**0.05**	**[0.04, 0.07]**	**0.03**	**[0.01, 0.04]**	**0.05**	**[0.04, 0.06]**	**0.04**	**[0.03, 0.05]**	**0.05**	**[0.04, 0.06]**
LISS	**-0.11**	**[-0.12, -0.10]**	**-0.09**	**[-0.10, -0.08]**	**-0.11**	**[-0.12, -0.10]**	**-0.09**	**[-0.10, -0.08]**	**-0.08**	**[-0.09, -0.07]**
(Age^2)	**0.05**	**[0.04, 0.06]**	**0.04**	**[0.03, 0.05]**	**0.05**	**[0.04, 0.06]**	**0.03**	**[0.02, 0.04]**	**0.04**	**[0.03, 0.05]**
NLSY	**-0.22**	**[-0.35, -0.09]**	**-0.19**	**[-0.31, -0.07]**	**-0.34**	**[-0.49, -0.18]**	-0.11	[-0.24, 0.01]	-0.09	[-0.21, 0.03]
(Age^2)	**0.22**	**[0.10, 0.35]**	**0.23**	**[0.10, 0.35]**	**0.30**	**[0.15, 0.46]**	0.13	[-0.00, 0.25]	0.12	[-0.00, 0.24]
Cognitive Ability										
GSOEP	**-0.07**	**[-0.08, -0.06]**	**-0.06**	**[-0.07, -0.05]**	**-0.05**	**[-0.06, -0.03]**	**-0.06**	**[-0.07, -0.04]**	**-0.10**	**[-0.11, -0.09]**
HILDA	**-0.08**	**[-0.09, -0.07]**	**-0.13**	**[-0.15, -0.12]**	**-0.10**	**[-0.11, -0.08]**	**-0.12**	**[-0.13, -0.11]**	**-0.16**	**[-0.17, -0.15]**
HRS	**-0.20**	**[-0.21, -0.18]**	**-0.20**	**[-0.22, -0.19]**	**-0.29**	**[-0.31, -0.28]**	**-0.18**	**[-0.19, -0.17]**	**-0.29**	**[-0.30, -0.27]**
NLSY	**-0.13**	**[-0.15, -0.12]**	**-0.18**	**[-0.20, -0.17]**	**-0.12**	**[-0.14, -0.10]**	**-0.16**	**[-0.17, -0.14]**	**-0.18**	**[-0.20, -0.16]**
Education Level										
GSOEP	**-0.15**	**[-0.17, -0.13]**	**-0.15**	**[-0.17, -0.13]**	**-0.08**	**[-0.10, -0.06]**	**-0.13**	**[-0.15, -0.11]**	**-0.23**	**[-0.25, -0.21]**
HILDA	**-0.09**	**[-0.11, -0.07]**	**-0.15**	**[-0.17, -0.12]**	**-0.09**	**[-0.11, -0.07]**	**-0.14**	**[-0.15, -0.12]**	**-0.18**	**[-0.20, -0.17]**
HRS	**-0.16**	**[-0.19, -0.14]**	**-0.10**	**[-0.13, -0.07]**	**-0.28**	**[-0.30, -0.25]**	**-0.15**	**[-0.17, -0.12]**	**-0.26**	**[-0.29, -0.24]**
LISS	**-0.07**	**[-0.09, -0.05]**	**-0.10**	**[-0.12, -0.08]**	**-0.07**	**[-0.09, -0.05]**	**-0.06**	**[-0.08, -0.04]**	**-0.03**	**[-0.05, -0.01]**
NLSY	**-0.19**	**[-0.22, -0.15]**	**-0.24**	**[-0.26, -0.21]**	**-0.20**	**[-0.24, -0.16]**	**-0.25**	**[-0.28, -0.22]**	**-0.19**	**[-0.23, -0.16]**
Income (log)										
GSOEP	**-0.05**	**[-0.05, -0.04]**	**-0.02**	**[-0.03, -0.01]**	**-0.08**	**[-0.09, -0.07]**	**-0.06**	**[-0.07, -0.05]**	**-0.06**	**[-0.06, -0.05]**
HILDA	**-0.05**	**[-0.06, -0.05]**	**-0.04**	**[-0.05, -0.04]**	**-0.05**	**[-0.06, -0.04]**	**-0.04**	**[-0.05, -0.03]**	**-0.05**	**[-0.06, -0.05]**

HRS	**-0.04**	**[-0.05, -0.03]**	-0.00	[-0.02, 0.01]	**-0.08**	**[-0.10, -0.07]**	**-0.04**	**[-0.05, -0.03]**	**-0.07**	**[-0.08, -0.05]**
LISS	**-0.02**	**[-0.02, 0.01]**	**-0.02**	**[-0.02, -0.01]**	**-0.02**	**[-0.03, -0.02]**	**-0.01**	**[-0.02, -0.01]**	**-0.02**	**[-0.02, -0.01]**
NLSY	**-0.02**	**[-0.03, -0.02]**	**-0.01**	**[-0.02, -0.01]**	**-0.03**	**[-0.03, -0.02]**	**-0.02**	**[-0.02, -0.02]**	**-0.02**	**[-0.02, -0.01]**

Note. Est = maximum a posteriori (MAP) estimate. CI = 95% credible interval. Bold values indicate that the credible intervals do not contain 0.00. The estimates for the age covariates are from the models including both linear and quadratic age.

3.5.6 Personality Traits

Extraversion. Typically, having one standard deviation higher-than-average levels of extraversion was associated with less within-person variability around the trajectories of all other Big Five traits (Table 8; Table S26). The average percent change in sigma was -8.82% for agreeableness, -6.11% for conscientiousness, -1.52% for neuroticism, and -3.62% for openness.

Agreeableness. For levels of agreeableness, the most consistent effects were found for conscientiousness (Table 8; Table S27). Negative effects emerged across all datasets such that having one standard deviation higher-than-average levels was associated with a sigma an average of 10.40% smaller than the model average (i.e., the sigma value for conscientiousness at average levels of agreeableness). Negative effects also emerged for neuroticism (-2.44%) and openness (-3.06%) in 3/5 datasets.

Conscientiousness. Generally, higher conscientiousness levels were negatively related to sigma values in all other traits (Table 8; Table S28). People with one standard deviation higher-than-average levels had a sigma that was smaller than their model-average sigma by 4.73% for extraversion, 10.23% for agreeableness, 3.56% for neuroticism, and 4.14% for openness.

Neuroticism. With the exception of one null effect, having higher-than-average levels of neuroticism was associated with greater within-person variability around trajectories of the other Big Five traits (Table 8; Table S29). The average percent change in sigma was 5.59% for extraversion, 9.33% for agreeableness, 12.43% for conscientiousness, and 4.14% for openness.

Openness. Associations were somewhat inconsistent for openness levels (Table 8; Table S30). Higher-than-average levels were associated with a smaller sigma in agreeableness for 4/5 datasets; in extraversion and conscientiousness for 3/5 datasets; and in neuroticism for 2/5 datasets. Effects in the opposite direction also sometimes emerged for neuroticism and conscientiousness. The average percent change in sigma was -5.40% for extraversion, -8.93% for agreeableness, -8.00% for conscientiousness, and -0.94% for neuroticism.

Summary. Neuroticism was the trait for which its average levels were most frequently associated with within-person variability around trajectories of the other Big Five traits, with these effects always indicating that higher levels were associated with larger sigma values in other traits. Then, for agreeableness and especially conscientiousness, the opposite effect was nearly always found and indicated that higher levels of these traits were associated with smaller sigma values in the other Big Five traits. Less consistent trends emerged for extraversion and openness, but their effects mostly mimicked those of agreeableness and conscientiousness. All effects were especially large in magnitude for HRS.

Table 8

Associations of the Big Five Traits as Covariates with Within-Person Residual Variability for Trajectories of the Other Big Five Traits

	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Covariate	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
Extraversion										
GSOEP	–	–	0.01	[0.00, 0.02]	**-0.08**	**[-0.09, -0.07]**	**0.04**	**[0.03, 0.04]**	**-0.02**	**[-0.03, -0.01]**
HILDA	–	–	**-0.05**	**[-0.06, -0.04]**	-0.00	[-0.01, 0.00]	**-0.03**	**[-0.04, -0.02]**	0.00	[-0.01, 0.01]
HRS	–	–	**-0.42**	**[-0.43, -0.40]**	**-0.20**	**[-0.21, -0.19]**	**-0.05**	**[-0.06, -0.04]**	**-0.09**	**[-1.00, -0.08]**
LISS	–	–	**-0.03**	**[-0.04, -0.02]**	**0.02**	**[0.01, 0.03]**	0.01	[-0.00, 0.02]	**0.02**	**[0.01, 0.02]**
NLSY	–	–	**-0.03**	**[-0.04, -0.02]**	**-0.07**	**[-0.08, -0.05]**	**-0.05**	**[-0.06, -0.03]**	**-0.10**	**[-0.11, -0.09]**
Agreeableness										
GSOEP	**0.01**	**[0.01, 0.02]**	–	–	**-0.16**	**[-0.17, -0.16]**	**0.04**	**[0.03, 0.05]**	**0.03**	**[0.02, 0.03]**
HILDA	**0.02**	**[0.01, 0.03]**	–	–	**-0.05**	**[-0.06, -0.04]**	**-0.09**	**[-0.10, -0.08]**	**-0.02**	**[-0.03, -0.01]**
HRS	**-0.14**	**[-0.15, -0.12]**	–	–	**-0.21**	**[-0.22, -0.20]**	**-0.03**	**[-0.04, -0.02]**	**-0.08**	**[-0.09, -0.07]**
LISS	**-0.02**	**[-0.03, -0.01]**	–	–	**-0.03**	**[-0.04, -0.02]**	**0.02**	**[0.01, 0.03]**	-0.01	[-0.02, 0.00]
NLSY	-0.02	[-0.03, -0.00]	–	–	**-0.11**	**[-0.13, -0.10]**	**-0.07**	**[-0.09, -0.06]**	**-0.08**	**[-0.09, -0.07]**
Conscientiousness										
GSOEP	-0.01	[-0.02, -0.00]	**-0.03**	**[-0.04, -0.03]**	–	–	**0.06**	**[0.05, 0.06]**	**0.04**	**[0.03, 0.05]**
HILDA	**-0.03**	**[-0.04, -0.02]**	**-0.11**	**[-0.12, -0.10]**	–	–	**-0.10**	**[-0.11, -0.09]**	**-0.04**	**[-0.05, -0.03]**
HRS	**-0.16**	**[-0.17, -0.15]**	**-0.37**	**[-0.39, -0.36]**	–	–	**-0.08**	**[-0.09, -0.07]**	**-0.14**	**[-0.15, -0.12]**
LISS	**-0.03**	**[-0.04, -0.02]**	**-0.05**	**[-0.06, -0.04]**	–	–	-0.00	[-0.01, 0.01]	**-0.02**	**[-0.03, -0.01]**
NLSY	**-0.02**	**[-0.04, -0.01]**	-0.02	[-0.03, -0.00]	–	–	**-0.07**	**[-0.09, -0.06]**	**-0.06**	**[-0.07, -0.04]**
Neuroticism										
GSOEP	**0.04**	**[0.04, 0.05]**	**0.03**	**[0.02, 0.03]**	**0.07**	**[0.06, 0.07]**	–	–	**0.02**	**[0.02, 0.03]**
HILDA	**0.03**	**[0.02, 0.04]**	**0.13**	**[0.12, 0.14]**	**0.11**	**[0.10, 0.12]**	–	–	-0.01	[-0.02, 0.00]
HRS	**0.11**	**[0.10, 0.12]**	**0.16**	**[0.15, 0.17]**	**0.16**	**[0.14, 0.17]**	–	–	**0.08**	**[0.06, 0.09]**
LISS	**0.05**	**[0.04, 0.06]**	**0.07**	**[0.05, 0.08]**	**0.06**	**[0.05, 0.07]**	–	–	**0.04**	**[0.03, 0.05]**
NLSY	**0.04**	**[0.02, 0.05]**	**0.05**	**[0.04, 0.07]**	**0.18**	**[0.16, 0.20]**	–	–	**0.07**	**[0.05, 0.08]**
Openness										
GSOEP	**-0.09**	**[-0.09, -0.08]**	**-0.04**	**[-0.05, -0.04]**	**-0.10**	**[-0.10, -0.09]**	-0.00	[-0.01, 0.01]	–	–
HILDA	0.01	[-0.00, 0.02]	**-0.05**	**[-0.06, -0.04]**	**0.02**	**[0.01, 0.02]**	**0.02**	**[0.02, 0.03]**	–	–
HRS	**-0.18**	**[-0.19, -0.17]**	**-0.37**	**[-0.38, -0.36]**	**-0.24**	**[-0.25, -0.23]**	**-0.05**	**[-0.06, -0.04]**	–	–
LISS	0.01	[0.00, 0.02]	-0.01	[-0.02, -0.00]	0.01	[-0.00, 0.02]	**0.03**	**[0.02, 0.04]**	–	–
NLSY	**-0.04**	**[-0.05, -0.02]**	**-0.04**	**[-0.05, -0.02]**	**-0.13**	**[-0.14, -0.11]**	**-0.05**	**[-0.06, -0.03]**	–	–

Note. Est = maximum a posteriori (MAP) estimate. CI = 95% credible interval. Bold values indicate that the credible intervals do not contain 0.00.

3.5.7 Life Events

For life events, the consistency of effects across datasets was often contingent upon the specific event (Tables 9-10; Tables S31-S36). Effects typically emerged far more frequently for the time-invariant, between-person variables than the time-varying, within-person variables. Between-person effects were both positive and negative in direction. Events associated with mature, adult roles and generally seen as positive typically had between-person effects that were negative in direction (i.e., associated with less within-person variability). In comparison, events associated with the loss of an adult role or that involve what is often considered a negative life experience typically had positive effects (i.e., associated with greater within-person variability). Then, meaningful within-person effects, without exception, were always positive in direction. The event-specific patterns that frequently emerged are described below.

Marriage. Compared to people that have never been married, individuals that reported ever being married had less within-person variability around their trajectories for extraversion and conscientiousness in 5/5 datasets and agreeableness in 4/5 datasets (Table 9; Table S31). Additionally, between-person effects emerged for all traits in both HILDA and NLSY, with these associations always being negative. There were no frequently emerging within-person effects.

For between-person effects, the average percent change in sigma for those who reported ever being married, relative to people never married, was -11.23% for extraversion, -9.32% for agreeableness, -14.97% for conscientiousness, -6.06% for neuroticism, and -3.80% for openness. For datasets, effects were largest in magnitude for LISS (-14.25%) and HILDA (-13.05%).

Divorce. Compared to people that have never been divorced, individuals that reported a divorce had greater within-person variability around their trajectory of neuroticism in 3/5 datasets (Table 9; Table S32). Then, the time-varying experience of getting divorced (i.e., the within-person effect) was associated with an increase in sigma for extraversion in 4/5 datasets. In LISS, getting divorced was also associated with increases in sigma for neuroticism and openness.

For between-person effects, the average percent change was generally small across traits and datasets, with conscientiousness (-4.09%) and HILDA (-3.72%) having the largest values. For within-person variables, effects were larger in magnitude and ranged from 2.05% (NLSY) to 13.19% (LISS) for datasets and from 3.05% (agreeableness) to 11.47% (extraversion) for traits.

Unemployment. Out of all life events, the most effects were found for unemployment (Table 9; Table S33). Individuals that reported ever being unemployed had greater within-person variability for trajectories of all Big Five traits compared to those who never reported this event. For time-varying effects, being unemployed was associated with increases in sigma for neuroticism in 5/5 datasets and for extraversion and conscientiousness in 3/5 datasets. This suggests that not only do individuals that have ever been unemployed have larger within-person variability around these trajectories, but further that the actual onset of unemployment is associated with an increase in their already larger degree of within-person variability.

For between-person effects, the average percent change in sigma for those who were ever unemployed, relative to those who were never unemployed, ranged from 8.56% (GSOEP) to 13.22% (LISS) for datasets and from 7.95% (agreeableness) to 14.43% (conscientiousness) for traits. For within-person variables, effects were generally smaller in magnitude and ranged from 2.85% (LISS) to 10.08% (HILDA) for datasets and from 3.95% (agreeableness) to 7.27% (neuroticism) for traits.

Table 9

Associations of Marriage, Divorce, and Unemployment as Covariates with Within-Person Residual Variability for Trajectories of the Big Five Traits

		Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Covariate	Term	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
Marriage											
GSOEP	B	**-0.07**	**[-0.09, -0.05]**	0.00	[-0.02, 0.02]	**-0.22**	**[-0.24, -0.20]**	-0.02	[-0.04, -0.00]	**0.04**	**[0.02, 0.06]**
	W	0.02	[-0.03, 0.06]	-0.00	[-0.04, 0.04]	-0.01	[-0.05, 0.02]	0.01	[-0.03, 0.05]	0.01	[-0.04, 0.05]
HILDA	B	**-0.13**	**[-0.15, -0.11]**	**-0.14**	**[-0.17, -0.12]**	**-0.15**	**[-0.17, -0.13]**	**-0.17**	**[-0.19, -0.15]**	**-0.11**	**[-0.14, -0.09]**
	W	0.02	[-0.02, 0.06]	-0.01	[-0.05, 0.04]	0.03	[-0.01, 0.07]	**0.06**	**[0.01, 0.10]**	0.01	[-0.03, 0.05]
HRS	B	**-0.07**	**[-0.12, -0.01]**	**-0.11**	**[-0.17, -0.05]**	**-0.06**	**[-0.12, -0.01]**	-0.03	[-0.08, 0.03]	-0.02	[-0.08, 0.03]
	W	-0.03	[-0.12, 0.05]	-0.03	[-0.09, 0.04]	-0.02	[-0.09, 0.06]	0.03	[-0.05, 0.11]	0.01	[-0.08, 0.09]
LISS	B	**-0.27**	**[-0.37, -0.17]**	**-0.19**	**[-0.28, -0.09]**	**-0.23**	**[-0.33, -0.14]**	-0.03	[-0.13, 0.08]	-0.07	[-0.17, 0.03]
	W	0.05	[-0.03, 0.12]	**0.10**	**[0.03, 0.17]**	0.02	[-0.05, 0.09]	0.06	[-0.01, 0.13]	0.05	[-0.02, 0.12]
NLSY	B	**-0.07**	**[-0.10, -0.04]**	**-0.06**	**[-0.09, -0.04]**	**-0.16**	**[-0.20, -0.13]**	**-0.07**	**[-0.09, -0.04]**	**-0.04**	**[-0.07, -0.01]**
	W	0.04	[-0.02, 0.10]	-0.03	[-0.07, 0.02]	0.03	[-0.02, 0.08]	0.01	[-0.05, 0.06]	0.03	[-0.02, 0.07]
Divorce											
GSOEP	B	0.00	[-0.02, 0.02]	**0.03**	**[0.01, 0.05]**	**-0.05**	**[-0.07, -0.03]**	**0.05**	**[0.03, 0.07]**	0.02	[-0.00, 0.04]
	W	**0.12**	**[0.06, 0.17]**	0.05	[-0.01, 0.10]	0.00	[-0.05, 0.05]	-0.01	[-0.07, 0.05]	**0.07**	**[0.01, 0.13]**
HILDA	B	-0.02	[-0.04, -0.00]	**-0.04**	**[-0.07, -0.02]**	**-0.06**	**[-0.08, -0.03]**	**-0.04**	**[-0.06, -0.01]**	**-0.03**	**[-0.06, -0.01]**
	W	**0.13**	**[0.06, 0.19]**	0.03	[-0.03, 0.10]	0.07	[0.00, 0.13]	0.02	[-0.04, 0.08]	0.01	[-0.05, 0.08]
HRS	B	0.00	[-0.02, 0.03]	**-0.05**	**[-0.08, -0.03]**	-0.01	[-0.03, 0.01]	**0.04**	**[0.02, 0.06]**	-0.01	[-0.03, 0.01]
	W	**0.11**	**[0.02, 0.20]**	0.02	[-0.06, 0.09]	0.02	[-0.06, 0.10]	**0.10**	**[0.01, 0.18]**	0.09	[0.00, 0.18]
LISS	B	-0.01	[-0.04, 0.02]	-0.01	[-0.04, 0.02]	-0.03	[-0.06, 0.00]	0.00	[-0.03, 0.04]	-0.01	[-0.04, 0.02]
	W	**0.14**	**[0.02, 0.27]**	0.03	[-0.09, 0.16]	0.08	[-0.04, 0.20]	**0.21**	**[0.09, 0.33]**	**0.15**	**[0.03, 0.27]**
NLSY	B	0.02	[-0.02, 0.07]	-0.01	[-0.06, 0.03]	-0.06	[-0.12, -0.00]	**0.06**	**[0.01, 0.11]**	0.05	[-0.00, 0.09]
	W	0.04	[-0.08, 0.15]	0.02	[-0.08, 0.12]	-0.01	[-0.12, 0.10]	0.05	[-0.05, 0.16]	-0.00	[-0.10, 0.10]
Unemployment											
GSOEP	B	**0.09**	**[0.07, 0.11]**	**0.08**	**[0.06, 0.10]**	**0.06**	**[0.04, 0.08]**	**0.08**	**[0.06, 0.10]**	**0.10**	**[0.09, 0.12]**
	W	**0.09**	**[0.04, 0.13]**	0.04	[-0.00, 0.08]	**0.04**	**[0.01, 0.08]**	**0.06**	**[0.02, 0.11]**	**0.06**	**[0.01, 0.10]**
HILDA	B	**0.10**	**[0.08, 0.12]**	**0.08**	**[0.06, 0.10]**	**0.13**	**[0.11, 0.15]**	**0.08**	**[0.07, 0.10]**	**0.11**	**[0.09, 0.13]**
	W	**0.10**	**[0.08, 0.13]**	**0.09**	**[0.07, 0.12]**	**0.09**	**[0.06, 0.12]**	**0.10**	**[0.07, 0.12]**	**0.10**	**[0.07, 0.12]**
HRS	B	**0.11**	**[0.09, 0.13]**	0.03	[0.00, 0.06]	**0.18**	**[0.15, 0.20]**	**0.13**	**[0.11, 0.16]**	**0.15**	**[0.13, 0.18]**
	W	**0.10**	**[0.05, 0.15]**	-0.02	[-0.06, 0.02]	0.03	[-0.01, 0.07]	**0.08**	**[0.03, 0.12]**	0.01	[-0.04, 0.07]
LISS	B	**0.13**	**[0.11, 0.15]**	**0.13**	**[0.10, 0.15]**	**0.14**	**[0.12, 0.16]**	**0.13**	**[0.11, 0.15]**	**0.09**	**[0.08, 0.11]**
	W	0.01	[-0.03, 0.05]	0.01	[-0.03, 0.05]	0.04	[0.00, 0.09]	**0.05**	**[0.01, 0.10]**	0.03	[-0.01, 0.07]

NLSY	B	**0.07**	**[0.02, 0.13]**	**0.06**	**[0.01, 0.11]**	**0.16**	**[0.09, 0.22]**	**0.09**	**[0.04, 0.14]**	**0.10**	**[0.04, 0.15]**
	W	0.03	[-0.01, 0.06]	**0.07**	**[0.05, 0.10]**	**0.07**	**[0.04, 0.10]**	**0.06**	**[0.03, 0.09]**	0.02	[-0.01, 0.05]

Note. B = between-person, time-invariant life event variable. W = within-person, time-varying life event variable. Est = maximum a posteriori (MAP) estimate. CI = 95% credible interval. Bold values indicate that the credible intervals do not contain 0.00.

Degree Attainment. Across all datasets, individuals with a university degree had less variability around their trajectories of all traits compared to individuals without a degree (Table 10; Table S34). Then, for time-varying effects, the experience of getting a degree was associated with increases in sigma for extraversion and neuroticism in 3/4 datasets and agreeableness and conscientiousness in 2/4 datasets. This suggests that, in general, people with less within-person variability tend to have university degrees, but the actual experience of obtaining the degree is, at least in the short-term, associated with increased within-person variability, regardless of the trait.

Effects for traits were typically larger for the between-person associations compared to the within-person associations. For between-person effects, the average percent change ranged from -6.37% (LISS) to -18.59% (NLSY) for datasets and from -10.53% (conscientiousness) to -14.97% (openness) for traits. For within-person variables, effects ranged from -2.00% (NLSY) to 14.00% (HILDA) for datasets and from 5.80% (openness) to 9.11% (neuroticism) for traits.

Child. Effects for having a child were somewhat inconsistent (Table 10; Table S35). For between-person effects, the more robust pattern was that people with children had smaller sigma values for extraversion and conscientiousness compared to people without children. For dataset-specific patterns, in HILDA and LISS, individuals with children had smaller sigma values for all Big Five traits relative to people without children. However, in NLSY, individuals that reported ever having a child had larger sigma values for agreeableness, neuroticism, and openness.

Then, the time-varying experience of this event was associated with increases in sigma for extraversion and neuroticism in 3/4 datasets and for conscientiousness in 2/4 datasets. This suggests individuals that have less within-person variability around their trajectories of most traits are perhaps more likely to have children – at least in datasets with an average age above young adulthood – but the actual experience of having a child is, at least temporarily, associated with increased within-person variability for extraversion, conscientiousness, and neuroticism.

For between-person effects, the average percent change ranged from -11.29% (LISS) to 5.03% (NLSY) for datasets and from -10.94% (conscientiousness) to 0.05% (openness) for traits. For within-person variables, effects ranged from 1.44% (NLSY) to 7.91% (LISS) for datasets and from 2.84% (agreeableness) to 5.97% (neuroticism) for traits.

Health Event. For health events, results were more consistent within datasets than they were across traits (Table 10; Table S36). In HRS, ever experiencing a health event was associated with larger sigma values for all Big Five traits. In GSOEP, similar effects emerged for agreeableness, neuroticism, and openness. In comparison, in LISS, people who ever reported a health event had less within-person variability for all traits except neuroticism. For time-varying effects, experiencing a health event was associated with increases in sigma for agreeableness, conscientiousness, and openness in LISS and for all traits except openness in HILDA.

Generally, effects were small in magnitude for both the between- and within-person variables. For between-person effects, the average percent change in sigma was largest in HRS (8.86%) for datasets and in neuroticism (2.04%) for traits. For within-person variables, the average percent change was largest in LISS (4.73%) and neuroticism (3.48%).

Table 10

Associations of Degree Attainment, Having a Child, and Experiencing a Health Event as Covariates with Within-Person Residual Variability for Trajectories of the Big Five Traits

		Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Covariate	Term	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
Degree Attainment											
GSOEP	B	**-0.16**	**[-0.18, -0.14]**	**-0.14**	**[-0.16, -0.12]**	**-0.08**	**[-0.10, -0.06]**	**-0.13**	**[-0.15, -0.11]**	**-0.23**	**[-0.26, -0.21]**
	W	**0.10**	**[0.03, 0.17]**	-0.01	[-0.08, 0.06]	0.05	[-0.01, 0.11]	**0.11**	**[0.05, 0.18]**	0.06	[-0.01, 0.13]
HILDA	B	**-0.11**	**[-0.13, -0.09]**	**-0.16**	**[-0.19, -0.14]**	**-0.12**	**[-0.14, -0.10]**	**-0.15**	**[-0.17, -0.13]**	**-0.20**	**[-0.22, -0.18]**
	W	**0.10**	**[0.05, 0.15]**	**0.14**	**[0.10, 0.19]**	**0.18**	**[0.14, 0.23]**	**0.13**	**[0.08, 0.18]**	**0.11**	**[0.07, 0.16]**
LISS	B	**-0.07**	**[-0.09, -0.05]**	**-0.10**	**[-0.12, -0.08]**	**-0.06**	**[-0.08, -0.04]**	**-0.06**	**[-0.08, -0.04]**	**-0.04**	**[-0.06, -0.02]**
	W	**0.09**	**[0.01, 0.16]**	**0.11**	**[0.03, 0.19]**	0.06	[-0.01, 0.14]	**0.13**	**[0.05, 0.20]**	**0.08**	**[0.01, 0.16]**
NLSY	B	**-0.18**	**[-0.21, -0.14]**	**-0.23**	**[-0.27, -0.20]**	**-0.19**	**[-0.23, -0.15]**	**-0.24**	**[-0.27, -0.20]**	**-0.19**	**[-0.22, -0.15]**
	W	0.00	[-0.07, 0.07]	0.01	[-0.05, 0.07]	-0.06	[-0.12, 0.00]	-0.03	[-0.09, 0.04]	-0.03	[-0.09, 0.03]
Child											
GSOEP	B	**-0.04**	**[-0.06, -0.03]**	0.01	[-0.01, 0.02]	**-0.19**	**[-0.21, -0.18]**	-0.02	[-0.03, 0.00]	**0.04**	**[0.03, 0.06]**
	W	**0.08**	**[0.04, 0.11]**	**0.06**	**[0.03, 0.09]**	0.02	[-0.00, 0.05]	**0.07**	**[0.04, 0.11]**	0.04	[0.00, 0.07]
HILDA	B	**-0.10**	**[-0.12, -0.08]**	**-0.09**	**[-0.11, -0.07]**	**-0.11**	**[-0.13, -0.09]**	**-0.10**	**[-0.12, -0.07]**	**-0.05**	**[-0.07, -0.03]**
	W	0.01	[-0.03, 0.05]	-0.00	[-0.04, 0.03]	**0.05**	**[0.01, 0.09]**	**0.04**	**[0.01, 0.08]**	0.02	[-0.01, 0.06]
LISS	B	**-0.14**	**[-0.16, -0.11]**	**-0.12**	**[-0.14, -0.10]**	**-0.14**	**[-0.17, -0.12]**	**-0.10**	**[-0.12, -0.08]**	**-0.10**	**[-0.12, -0.07]**
	W	**0.08**	**[0.02, 0.15]**	0.06	[-0.01, 0.12]	**0.09**	**[0.03, 0.15]**	**0.10**	**[0.04, 0.16]**	0.05	[-0.01, 0.12]
NLSY	B	0.02	[-0.01, 0.05]	**0.08**	**[0.05, 0.11]**	-0.03	[-0.06, 0.01]	**0.07**	**[0.04, 0.10]**	**0.10**	**[0.07, 0.13]**
	W	**0.06**	**[0.01, 0.10]**	-0.01	[-0.05, 0.03]	-0.01	[-0.05, 0.04]	0.02	[-0.02, 0.06]	0.01	[-0.03, 0.05]
Health Event											
GSOEP	B	0.01	[-0.00, 0.03]	**0.04**	**[0.02, 0.05]**	0.00	[-0.02, 0.02]	**0.04**	**[0.03, 0.06]**	**0.04**	**[0.02, 0.06]**
	W	0.02	[-0.01, 0.04]	0.02	[-0.00, 0.05]	0.01	[-0.02, 0.03]	0.03	[-0.00, 0.05]	0.00	[-0.03, 0.03]
HILDA	B	**0.04**	**[0.02, 0.06]**	-0.00	[-0.02, 0.02]	0.01	[-0.01, 0.03]	-0.00	[-0.02, 0.02]	0.01	[-0.01, 0.03]
	W	**0.04**	**[0.01, 0.07]**	**0.04**	**[0.01, 0.07]**	**0.05**	**[0.03, 0.08]**	**0.07**	**[0.04, 0.09]**	0.03	[0.00, 0.05]
HRS	B	**0.09**	**[0.06, 0.11]**	**0.06**	**[0.03, 0.09]**	**0.16**	**[0.13, 0.19]**	**0.04**	**[0.02, 0.07]**	**0.07**	**[0.05, 0.10]**
	W	0.01	[-0.02, 0.03]	-0.02	[-0.04, 0.00]	-0.02	[-0.05, 0.00]	0.03	[-0.00, 0.05]	0.01	[-0.02, 0.03]
LISS	B	**-0.06**	**[-0.08, -0.03]**	**-0.06**	**[-0.08, -0.04]**	**-0.08**	**[-0.11, -0.06]**	-0.01	[-0.04, 0.01]	**-0.06**	**[-0.08, -0.03]**
	W	0.02	[-0.02, 0.06]	**0.05**	**[0.01, 0.08]**	**0.04**	**[0.01, 0.08]**	0.04	[0.00, 0.08]	**0.08**	**[0.05, 0.12]**
NLSY	B	-0.05	[-0.09, -0.00]	**-0.06**	**[-0.10, -0.02]**	-0.04	[-0.09, 0.01]	0.03	[-0.01, 0.07]	-0.03	[-0.08, 0.01]
	W	-0.02	[-0.05, 0.02]	-0.01	[-0.04, 0.02]	0.03	[-0.01, 0.06]	-0.00	[-0.03, 0.03]	-0.02	[-0.05, 0.02]

Note. B = between-person, time-invariant life event variable. W = within-person, time-varying life event variable. Est = maximum a posteriori (MAP) estimate. CI = 95% credible interval. Bold values indicate that the credible intervals do not contain 0.00.

3.6 Can Individual Differences in Sigma Predict an Outcome?

Lastly, I performed a test of the predictive utility of the person-level sigma values. Generally, this individual difference was negatively and robustly associated with health status (Table 11; Figures 10-11). Moreover, individual differences in sigma provided predictive utility above and beyond individuals' intercept and slope values. Many interaction effects for sigma with intercepts and slopes were found as well. Below, general trends and meta-analytic estimates across the three types of regression models are discussed. As a reminder, all effects are in correlation units, which permits the direct comparison of their magnitudes. Results from individual models across all datasets can be found in Tables S38-S39 and Figures S6-S8.

3.6.1 Simple Regressions

For simple regressions, sigma was negatively associated with health status for all Big Five traits, such that having larger-than-average sigma values was associated with below-average health status (Table 11; Table S38; Figure S6). Associations were usually largest for neuroticism and conscientiousness, and these traits had meta-analytic effects of -0.16 and -0.15, respectively.

3.6.2 Multiple Regressions

When controlling for levels and changes in traits, individual differences in sigma proved to still have predictive utility (Table 11; Table S38; Figure S7). Sigma was negatively associated with health status for conscientiousness, neuroticism, and openness in 4/5 datasets and for extraversion and agreeableness in 3/5 datasets. Meta-analytic effects were largest in magnitude for conscientiousness (-0.10) and openness (-0.08). The average magnitude of the effects for sigma was equivalent to that of slopes, with an average of |0.07|. Intercepts had an average magnitude that was double that of sigma and slope values, with a value of |0.14|.

Table 11

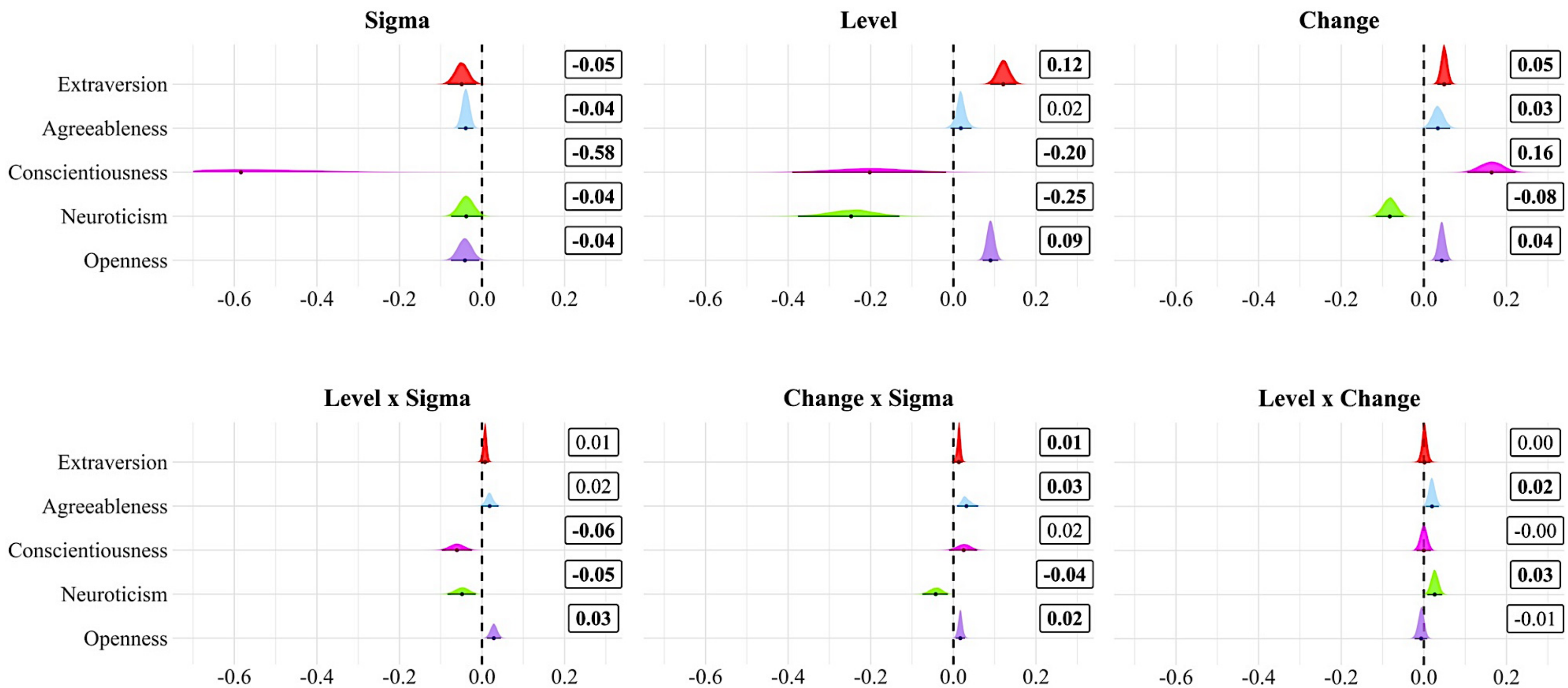
Meta-Analytic Estimates from the Regression Models for Predicting Health Status

		Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
Model	Term	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
Simple											
	Sigma	**-0.11**	**[-0.17, -0.04]**	**-0.09**	**[-0.12, -0.06]**	**-0.15**	**[-0.19, -0.11]**	**-0.16**	**[-0.19, -0.12]**	**-0.10**	**[-0.17, -0.03]**
Multiple											
	Sigma	**-0.06**	**[-0.09, -0.02]**	**-0.06**	**[-0.09, -0.03]**	**-0.10**	**[-0.14, -0.05]**	**-0.05**	**[-0.09, -0.02]**	**-0.08**	**[-0.12, -0.03]**
	Level	**0.12**	**[0.07, 0.18]**	**-0.15**	**[-0.25, -0.06]**	**0.20**	**[0.02, 0.39]**	**-0.19**	**[-0.27, -0.12]**	**0.05**	**[0.01, 0.09]**
	Change	**0.05**	**[0.03, 0.07]**	0.02	[-0.02, 0.06]	**0.16**	**[0.07, 0.25]**	**-0.07**	**[-0.08, -0.05]**	**0.05**	**[0.03, 0.07]**
Interaction											
	Sigma	**-0.05**	**[-0.08, -0.01]**	**-0.04**	**[-0.06, -0.02]**	**-0.58**	**[-0.93, -0.25]**	-0.04	[-0.07, -0.00]	**-0.04**	**[-0.07, -0.01]**
	Level	**0.12**	**[0.09, 0.15]**	0.02	[-0.01, 0.04]	**-0.20**	**[-0.39, -0.02]**	**-0.25**	**[-0.38, -0.13]**	**0.09**	**[0.07, 0.11]**
	Change	**0.05**	**[0.03, 0.07]**	**0.03**	**[0.01, 0.06]**	**0.16**	**[0.11, 0.22]**	**-0.08**	**[-0.12, -0.05]**	**0.04**	**[0.03, 0.06]**
	L x S	0.01	[-0.00, 0.02]	0.02	[-0.00, 0.04]	**-0.06**	**[-0.10, -0.02]**	**-0.05**	**[-0.08, -0.02]**	**0.03**	**[0.01, 0.05]**
	C x S	0.01	[0.00, 0.02]	**0.03**	**[0.01, 0.06]**	0.02	[-0.01, 0.06]	**-0.04**	**[-0.07, -0.01]**	**0.02**	**[0.01, 0.03]**
	L x C	0.00	[-0.01, 0.02]	0.02	[0.00, 0.04]	0.00	[-0.02, 0.02]	**0.03**	**[0.01, 0.04]**	-0.01	[-0.02, 0.01]

Note. Est = maximum a posteriori (MAP) estimate. CI = 95% credible interval. Sigma (S) = person-level sigma values. Level (L) = person-level intercept (i.e., initial trait level) values. Change (C) = person-level slope values. Bold values indicate that the credible intervals do not contain 0.00.

Figure 10

Posterior Distributions of the Meta-Analytic Estimates for Main Effects and Interactions of Sigma, Intercept, and Slope Values Predicting Health Status



Note. Results are from the full interaction regression models. The horizontal lines for each effect delineate the 95% credible interval bounds and the solid dot indicates where the maximum a posteriori (MAP) estimate is. The vertical dashed, black line marks 0.00 on the x-axis. All estimates are in correlation units. Bolded values are those in which the 95% credible intervals for an effect did not include .00.

3.6.3 Interaction Effects

Lastly, a holistic picture emerged when examining interactions between sigma, intercept, and slope values (Table 11; Table S38; Figures 10-11; Figures S8-S17). For sigma, meta-analytic interaction effects with intercepts emerged for 3/5 traits and with slopes for 4/5 traits. The typical finding was that having larger sigma values would exacerbate a negative association that intercepts or slopes had with health status, such that even lower ratings were predicted. Having larger sigma values would also often magnify positive associations, such that having a large sigma and a beneficial intercept or slope would predict even better health status.

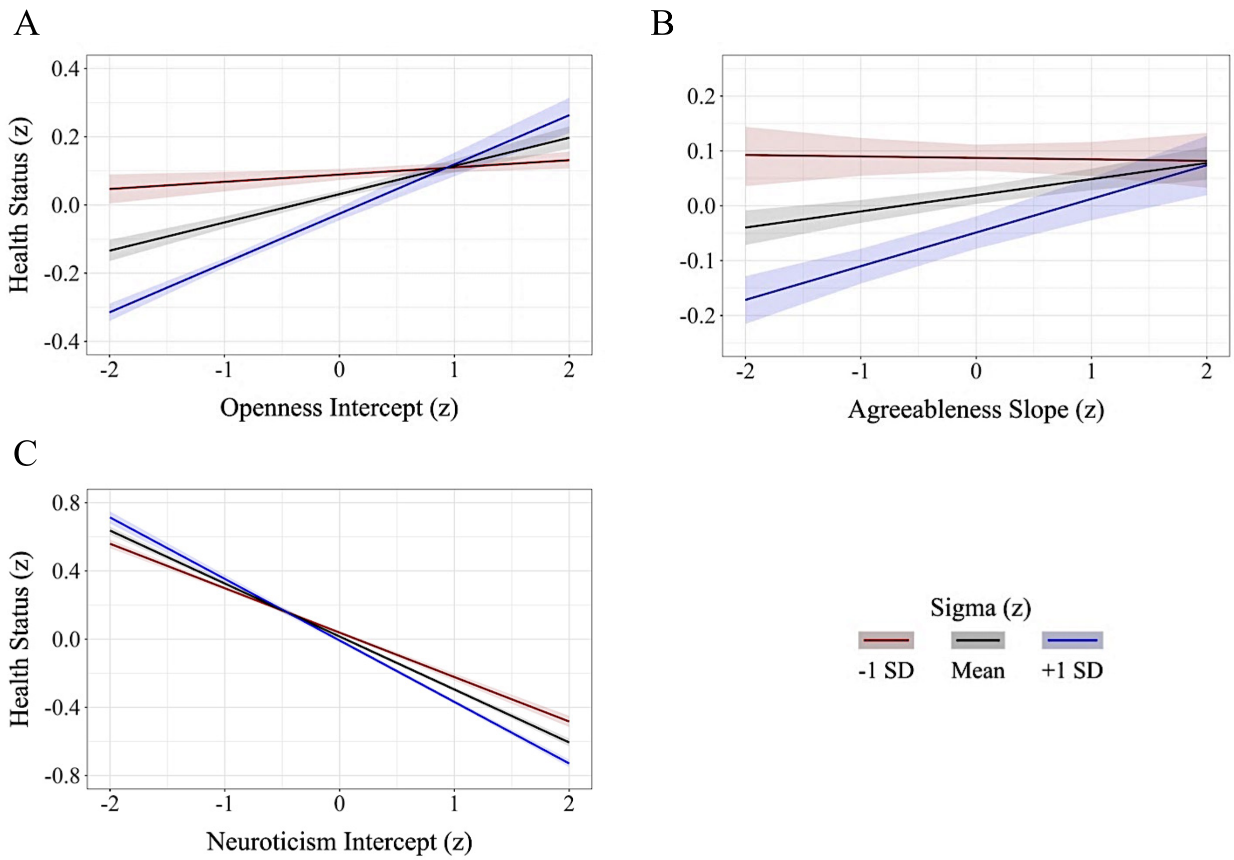
For instance, for majority of sigma-intercept interactions, effects were most evident at detrimental values of intercepts (i.e., predicted lower health status). For example, lower initial levels and larger sigma values for openness are negatively associated with health status. Their interaction thus predicted even lower ratings of health status for individuals with both low initial levels and large sigma values compared to those with smaller sigma values (Figure 11A).

A similar pattern emerged for sigma-slope interactions. For models in which sigma had a negative main effect, having slopes that are negatively associated with health, such as decreases in agreeableness, as well as larger sigma values predicted even lower health status (Figure 11B). At beneficial slope values, though, health status ratings are similar across different sigma values.

Additionally, when sigma has negative main effects, as it did in the previous examples, people's health status is usually similar regardless of their sigma value if they have beneficial intercept or slope values (i.e., those that predict higher health status). However, sometimes sigma had positive main effects in individual models. In these instances, sigma now also exacerbated effects at *beneficial* values of intercepts or slopes. For example, having beneficial levels of neuroticism (i.e., low) and a large sigma value predicts even higher health status (Figure 11C).

Figure 11

Interactions of Levels and Changes in Traits with Sigma Values for Predicting Health Status



Note. Health status, intercepts, slopes, and sigma values are standardized (z).

3.6.4 Summary

Overall, meta-analytic effects indicated that having larger sigma values was negatively associated with health status for all Big Five traits. This was true even when controlling for initial trait levels and changes in traits, highlighting the robust predictive utility of sigma. Then, notably, the interaction models revealed that larger sigma values nearly always exacerbated the negative associations, and also frequently magnified the positive associations, that intercepts and slopes had with health status.

Chapter 4: Discussion

In this study, I examined the degree to which there were individual differences in within-person variability around trajectories of the Big Five, if a comprehensive set of variables central to personality development were associated with this heterogeneity, and if the person-level sigma values could predict an outcome, above and beyond the effects of trait levels and changes in traits. Across all models, there were meaningful individual differences in longitudinal within-person variability – the magnitude of which was similar to, and frequently even greater than, that of individual differences in levels and changes in the same traits. Moreover, there was never a case in which a MELSM did not have a substantially better model fit than its corresponding standard MLM, which further underscores the shortcomings in typical modeling approaches. Variables routinely shown to be important for other aspects of personality development were also meaningfully associated with individual differences in sigma, suggesting this individual difference can similarly be leveraged to understand the complex ways in which personality develops and interacts with the environment. Lastly, the person-level sigma values proved to not only have robust predictive utility, but they further offered unique insight into how personality may relate to important outcomes. Insight that, notably, trait levels and changes in traits cannot themselves equally provide. I discuss these results below, highlighting the nature and degree of individual differences in sigma; the implications these have for remedying issues and providing answers to unsolved questions in personality psychology; and the importance of empirically quantifying and theoretically incorporating individual differences in longitudinal within-person variability in future personality development research.

4.1 People Differ in Their Degree of Within-Person Variability for Trajectories of the Big Five Traits

Previously, consideration of an individual's own variability in a psychological construct has been confined to research examining within-person dynamics or changes in behavior over short time periods. In personality psychology, even definitions or descriptions of within-person variability often include words or phrases such as "fluctuations" (e.g., Howell et al., 2017; McCabe, 2017; Vazire & Sherman, 2017), "situations" (e.g., Baird et al., 2006; Geukes et al., 2017), "states" (Geukes et al., 2017; Vazire & Sherman, 2017), "short-term" (Nesselroade, 2001), and "very short time-periods" (Mõttus et al., 2020). These either imply or explicitly emphasize a certain temporal brevity to the psychological process or behavior being studied. Consequently, it may give the impression that similar applications over longer periods of time are not immediately compatible with existing work. Furthermore, both historical and modern discussions of within-person variability in empirical or theoretical articles are almost entirely in the context of state manifestations of traits and cross-situational consistency in behaviors from experience sampling studies (e.g., Beck & Jackson, 2021; Cattell, 1943; Cattell et al., 1947; Conner et al., 2009; Eid & Diener, 1999; Fleeson, 2001, 2004, 2012; Geukes et al., 2017; Heller et al., 2007; Magee et al., 2018; Mõttus et al., 2020; Ram & Gerstorf, 2009; Vazire & Sherman, 2017; Woodrow, 1932).

In comparison, for longitudinal research, the notion of variability within a person across longer periods of time (e.g., years) is often conceptualized as ipsative or personality profile consistency (Asendorpf & van Aken, 1991; De Fruyt et al., 2006; Donnellan et al., 2007; Ibáñez et al., 2016; Jackson & Beck, 2021; Ozer & Gjerde, 1989; Robins et al., 2001; Terracciano et al., 2010; Wright & Jackson, 2023b). This informs of how variable or (in)consistent someone is in the rank-order of their pattern of responding for trait indicators across two time points, often quantified with a profile correlation. Notably, other metrics conceptually related to profile correlations, such as the *D* indices proposed by Cronbach and Gleser (1953), have similar calculations and provide comparable information to some metrics used to quantify within-person variability (e.g., mean square successive difference; Jahng et al., 2008) – highlighting the conceptual and empirical links between within-person variability across short- and long-time scales. Unfortunately, though, profile consistency offers a narrow view of within-person variability and could easily be, and is often, rather considered a way to conceptualize personality change (Roberts et al., 2008). Two indices of between-person variability, namely variability in trait levels (i.e., intercepts) and changes in traits (i.e., slopes), are instead typically the focus in longitudinal personality development research that examines variability. However, the findings from the current study make a considerable case for future research to make it commonplace to additionally quantify individual differences in sigma.

Importantly, quantifying individual differences in sigma provides information about both between- and within-person variability. At the within-person level, the person-specific sigma values themselves indicate the degree to which people unsystematically vary in their repeated assessments across time versus remain quite consistent in their pattern of responding for a single construct. Accordingly, this quantifies the person-specific degree of the spread of residuals around one's line of best fit, or the amount of within-person variability around their predicted trajectory. This can be informative for understanding their true pattern of change and the mechanisms that may be underlying this change. For instance, if someone decreases in their degree of residual variability across time, such as in response to some external factor or simply passive development, then it is likely they are in an environment or had an experience that complements or serves to stabilize their personality, thus allowing for predictable patterns of personality development. In comparison, for someone that increased in their degree of within-person variability, this can perhaps be an indicator of shifting environmental factors (Denissen et al., 2019; Wright & Jackson, 2023c, 2023e), changes in attributes such as cognitive ability (Lönnqvist et al., 2008; Terracciano et al., 2018; Toomela, 2003), or reflect qualities such as personality inconsistency (Wright & Jackson, 2023b). For instance, in the present study, greater within-person variability around Big Five trajectories was associated with the time-varying experience of life events, lower cognitive ability, and personality traits linked to lower personality profile consistency (Donnellan et al., 2007; Wright & Jackson, 2023b), above and beyond any effects rather due to trait levels or changes in traits. Thus, this variability offers a valuable source of information about intraindividual personality development that could not equivalently be obtained by another individual difference variable.

Then, at the between-person level, individual differences in sigma indicate which individuals possess more or less of this within-person variability around their line of best fit. In the current study, dispersion metrics that quantified the degree of heterogeneity in the person-specific sigma values were calculated. When using these metrics to compare the degree of individual differences in sigma with that of individual differences in intercepts and slopes, a clear set of conclusions emerged. First, when comparing coefficients of variation for the person-level estimates, the degree of variability in sigma values was, on average, larger than it was for intercept values. Second, when converting all three individual difference metrics onto the same measurement scale, the average magnitude of the standard deviations for person-specific sigma values was larger than the magnitude of the standard deviations for the total amount of trait change people were predicted to have across time. Importantly, these findings highlight the degree of between-person heterogeneity present in longitudinal within-person variability for trajectories of the Big Five traits. They additionally contextualize the amount of between-person variability in sigma values by comparing it to the variability of the two most commonly examined individual differences in personality development – ultimately finding that the magnitude of individual differences in sigma are comparable to that of individual differences in initial trait levels and total changes in traits across time.

Overall, individual differences in within-person variability – a metric of personality that is typically reserved for use in investigations of short-term, dynamic processes – proved to similarly be a valuable source of information in a longitudinal examination of personality development spanning multiple years. What is often assumed to be random noise with no implications, and accordingly not even permitted to vary across individuals in typical models, was indeed found to be a meaningful source of individual differences. This finding replicated across all Big Five traits in each of the five datasets. When considering both standardized and unstandardized measures of dispersion, the degree of variability in the sigma values was, on average, larger than what was found for both intercepts and slopes – parameters that are regularly permitted to vary across individuals in longitudinal research. Regularly incorporating individual differences in longitudinal within-person residual variability into personality development research will serve to strengthen future work by quantifying another source of meaningful variance and expanding the repertoire of tools available for conducting this work.

4.2 Individual Differences in Within-Person Variability Can Help Solve Pieces of the Personality Development Puzzle

4.2.1 The Role and State of Description, Prediction, and Explanation in Personality Development

Notably, integrating individual differences in longitudinal within-person variability into research can help unify the three central goals of psychology – description, prediction, and explanation (Hamaker et al., 2020; Mõttus et al., 2020; Yarkoni & Westfall, 2017). The three goals are philosophically and conceptually compatible with each other, but empirically there often exist practical and statistical barriers to their harmonious integration (Yarkoni & Westfall, 2017). Somewhat paradoxically, it is simultaneously true that research motivated by one of the three goals is a) often isolated from the other forms, thus exacerbating their disjunction and b) sometimes poorly distinguished from the other goals (Shmueli, 2010; Yarkoni & Westfall, 2017), thus leading to ambiguity in the objectives and implications of a study (Hamaker et al., 2020). There are three goals for a reason, though: work motivated by one goal provides conclusions that research motivated by a different goal cannot equally provide. Similarly, one type of research can suffer from drawbacks and limitations that another is well-suited to overcome (Hamaker et al., 2020). Accordingly, it is vital to conduct all three forms of research – a comprehensive and cumulative body of knowledge can only be obtained through their (balanced) synthesis (Baumert et al., 2017).

However, the current state of the three goals and their balanced synthesis for personality development research is less than ideal, such that many pieces in the personality puzzle remain unsolved – or perhaps even missing. Despite being the predominant motivation for majority of research (Shmueli, 2010; Yarkoni & Westfall, 2017), (satisfactory) explanation remains as the most elusive of the three goals, likely owing to loosely applied theories and atheoretical, exploratory research practices that often guide this work (i.e., ad hoc revisions that reflect the concept of holist undetermination; Popper, 1959; Quine, 1980). Additionally, while there is no shortage of prediction research (e.g., Beck & Jackson, 2022; Soto, 2021; Wright & Jackson, 2022a, 2023a), its effects, although frequently present, often explain marginal amounts of variance and still have poorly understood underlying mechanisms (Götz et al., 2022; Mõttus et al., 2020). Then, description is the foundation of all research, but its value is nevertheless routinely discounted and purely descriptive work is the least frequently conducted (Hernán et al., 2019; Mõttus et al., 2020).

When these issues persist over time, much as they have for personality development research, it suggests they may emanate from the norms surrounding how research is conducted. This would accordingly call for alternative solutions that differ from the customary ways of carrying out research for each of the three goals. Thus far, proposed solutions typically advocate for an imbalanced conducting of the different types of research (Mõttus et al., 2020; Yarkoni & Westfall, 2017) and subsequently amplify their existing disconnect. For example, some have suggested that in order to meaningfully advance knowledge, strengthen empirical studies to optimize tests of interest, and improve subsequent research, researchers must choose one goal (Yarkoni & Westfall, 2017). Unfortunately, this is often only beneficial for work abiding by the same goal and can result in a harmful disregard of potential implications for other areas of research. Another recommendation is to focus on prediction, but to do so by developing predictive models that leverage many low-dimensional features (i.e., those with little to no aggregation) that optimize out-of-sample accuracy (e.g., Mõttus et al., 2020; Yarkoni & Westfall, 2017). However, not only would this exacerbate the imbalance in research supporting each of the three goals, but it may create the illusion that a specific set of analytic tools belongs exclusively to one type of research or, conversely, that research focused on one goal is restricted to a single methodological framework. Overall, encouraging restriction in research, whether it is restriction to a single goal or type of analysis, or both, does not seem to be a constructive course of action.

Instead, a viable solution for improving and advancing research motivated by a single goal, as well as integrating and harmonizing all research, is to continue expanding the realm of meaningful individual difference variables. Importantly, in any study, measured variables and the constructs they represent are agnostic to any one goal. For example, neuroticism is useful for a) description, such that its mean levels tend to decline across the lifespan (Bleidorn et al., 2022; Roberts et al., 2006) and women often score higher in this trait than men (Lehmann et al., 2013); b) prediction, such that it is robustly and prospectively associated with outcomes such as mortality, degree attainment, unemployment, salary, incarceration, and divorce (Converse et al., 2018; Mroczek & Spiro, 2007; Pusch et al., 2019; Wright & Jackson, 2022a, 2023a); and c) explanation, such that neuroticism itself is sometimes a mediator (e.g., Quilty et al., 2008) and its effects on future outcomes is sometimes mediated by other variables (e.g., Wright et al., 2022). The utility of a single individual difference variable within multiple domains of research serves to harmonize inferences, strengthen theories, and bridge gaps by capitalizing on the cumulative body of knowledge for that construct. One such variable with great potential for remedying existing issues and contributing to personality development research – that has notably already demonstrated it has descriptive value, predictive utility, and associations with a multitude of explanatory variables central to personality development – is individual differences in longitudinal within-person variability.

4.2.2 On Description and Individual Differences in Longitudinal Within-Person Variability

It would be an understatement to say that modern personality psychology would not exist without descriptive research. This is especially true considering the field is currently dominated by the trait paradigm, whose models were discovered and validated through descriptive work. Moreover, predictive and explanatory research cannot exist without descriptive work; some element of description is necessary for and precedes both prediction and explanation (Baumert et al., 2017; Eysenck & Eysenck, 2013). That is, description is the foundation of all research; it is necessary, but not sufficient, for a study with the goal of adding to the cumulative science of personality to be able to define, describe, and conceptually relate its variables of interest and their hypothesized associations. Unfortunately, the amount of new, purely descriptive research often pales in comparison to other research, likely owing to its negative connotations (e.g., lack of methodological/statistical rigor; Hernán et al., 2019; Mõttus et al., 2020). This has resulted in a boom in predictive and explanatory research attempting to rely on past descriptive work, which is often comparatively outdated and focused on description of mean levels and rank-order stability, to serve as the foundation for their studies. However, there remains much to be discovered beyond these common metrics of personality development.

When considering other ways to conceptualize personality, namely those beyond mean levels and rank-order stability, additional valuable descriptive information indeed exists. For instance, research using the Big Five has discovered what average levels of profile or ipsative consistency are (e.g., Donnellan et al., 2007; Terracciano et al., 2010; Wright & Jackson, 2023b), how one's level of personality consistency is itself a stable individual difference (Wright & Jackson, 2023b), how personality structure changes with age (e.g., Beck et al., 2022), individual differences in state personality distributions (e.g., Fleeson, 2001); the existence of individual differences in situation characteristic-state contingences (e.g., Kuper et al., 2022), and the stability and heterogeneous structure of idiographic personality networks (e.g., Beck & Jackson, 2020; Wright et al., 2019). Importantly, though, while these many ways of studying personality are often considered and treated as conceptually and empirically distinct from one another, they all simultaneously coexist and serve the same purpose of understanding personality. As such, it seems sensible to comprehensively describe the many possible individual differences that can be identified and assessed. Then, this knowledge can be used as foundational work for future research, help to better understand the many ways in which personality manifests and influences one's life, create the opportunity to revise pre-existing and formulate new theories, and allow these metrics to be used in service of predictive or explanatory research.

With regard to individual differences in sigma specifically, there are many great implications. Once foundational knowledge is acquired for a given individual difference variable, such as longitudinal within-person variability, it allows for its incorporation into other forms of research. Accordingly, a comprehensive body of knowledge can be obtained and then a better, more holistic understanding of personality development naturally follows. For example, this has already occurred in an area of research that utilizes within-person variability. This type of variability is not a novel concept and has been discussed in psychology for several decades (e.g., Allport, 1937; Cattell, 1943; Cattell et al., 1947; Woodrow, 1932). For personality, within-person variability was traditionally quantified using within-person or intraindividual standard deviations (Eid & Diener, 1999; Fleeson, 2001; Murray et al., 2002; Penner et al., 1994). Past work indeed found meaningful variability when quantified in this manner as well as between-person differences in the magnitude of this variability (e.g., Eid & Diener, 1999; Fleeson, 2001).

These findings then paved the way for descriptive work documenting its associations with other individual differences (e.g., Fleeson, 2001, 2007), prediction research leveraging it as both a predictor and outcome (e.g., Baird et al., 2006; Hardy & Segerstrom, 2017; Lievens et al., 2018; Ram & Gerstorf, 2009), and explanatory research examining its underlying mechanisms and reasons as to why it differs across individuals (e.g., Berry & Jobe, 2002; Geukes et al., 2017). As research accumulated, theories could be developed and continually refined to incorporate this variability as a fundamental individual difference that is crucial for understanding psychological processes (Berry & Jobe, 2002; Blum et al., 2018; Cervone, 2005; Hooker, 2002; Mischel & Shoda, 1995). Similarly, new methods and analytical frameworks were created (e.g., Blozis, 2022; Hedeker et al., 2008; Nestler, 2022; Ram et al., 2012) to address the increasingly complex questions researchers were able to ask because of the excellent foundational work in this area.

In sum, overlooking valuable sources of basic descriptive information, such as individual differences in within-person residual variability in longitudinal research, does a disservice to the field and creates opportunities for missing links in knowledge to be pervasive throughout research. In order to holistically understand personality development, research must start at and exhaust the simplest levels of description. Within-person variability has proved to be a fruitful topic for research examining personality dynamics and short-term processes; the same could be expected to happen for personality development.

4.2.3 On Prediction and Individual Differences in Longitudinal Within-Person Variability

There is extant research demonstrating the predictive utility of personality for important life outcomes (e.g., Beck & Jackson, 2022; Mroczek & Spiro, 2007; Roberts et al., 2007; Soto, 2021; Wright et al., 2022; Wright & Jackson, 2022a, 2023a). The ability of personality to predict outcomes exists for many metrics, including trait levels (Beck & Jackson, 2022; Wright & Jackson, 2023a), changes in traits (Mroczek & Spiro, 2007; Wright & Jackson, 2023a), profile consistency and similarity (Klimstra et al., 2010), and intraindividual variability across short time periods (Baird et al., 2006; Hardy & Segerstrom, 2017; Lievens et al., 2018; Ram & Gerstorf, 2009). Although it would be a folly to wholly discount the predictive utility of personality, in practice, this prediction is never perfect, less-than-ideal portions of variance are explained, and the magnitude of the effect sizes for these associations are generally small to medium (Götz et al., 2022). Accordingly, it is worth investigating other ways in which prediction can be optimized. One such way is leveraging other individual difference metrics for personality.

In the present study, even with only one test outcome, individual differences in sigma proved to have predictive utility above and beyond any associations attributable to initial levels and changes in traits. Additionally, a more complete picture of how sigma was related to one's health status emerged in the full interaction models. A few patterns are worth noting. First, the most robust pattern was that greater within-person residual variability in a trait often further worsens the effect of having a "negative" quality in a trait with respect to health status. That is, if high levels of a trait are negatively associated with health status, then also having a larger sigma value will predict even lower ratings of health status. This was generally true for levels and changes in all Big Five traits that had interaction effects. Second, it was also often the case that greater within-person residual variability conversely further exacerbated the effect of having a "positive" quality in a trait. This was true for associations of agreeableness, neuroticism, and openness with health status. Third, smaller sigma values for a trait sometimes served to attenuate or nullify negative associations that levels or changes in a trait had with health status, and in some cases even made them positive. This attenuating and/or nullifying of negative associations occurred for all Big Five traits and instances of making a negative association positive occurred for conscientiousness, neuroticism, and openness.

The many nuanced effects in the current study, as well as in any study in which some personality metric is used to predict an outcome, highlight an important point. To obtain the most accurate conclusions, researchers must leverage multiple indices and conceptualizations of personality. For instance, even when considering the two most common metrics – mean levels and changes in traits – it is quite often the case that their associations with an outcome differ in magnitude (e.g., Hoff et al., 2021; Mroczek & Spiro, 2007; Takahashi et al., 2013; Turiano, Pitzer, et al., 2012; Wright & Jackson, 2023a) and sometimes are even in opposite directions of one another (Wright & Jackson, 2023a). These patterns emerged in the present study as well. Additionally, it is evident that the effects one metric of personality has with an outcome are sometimes contingent upon another metric (i.e., interactions). Although past work has found that trait-by-trait interactions for mean levels occur relatively infrequently (Vize et al., 2022; Wright & Jackson, 2023f), this indicates little to nothing about the prevalence of interactions between different ways of quantifying personality (e.g., levels, changes, residual variability). Each aspect of personality does not exist in a black box; it coexists with the other aspects of one's entire personality and, so it seems, then differentially manifests in terms of how one behaves and navigates through life. Consequently, individuals' personalities, when considered in whole, have many unique associations with outcomes that will be overlooked, misquantified, or inaccurately attributed to another metric if these many important individual differences are not utilized. The current study suggests that incorporating individual differences in longitudinal within-person variability into prediction research for personality development would be a fruitful endeavor.

4.2.4 On Explanation and Individual Differences in Longitudinal Within-Person Variability

Research concerned with the remaining goal, explanation, would similarly benefit from integrating individual differences in sigma as a core construct. In psychology, it is often believed that explanatory research is reserved for studies that conduct experiments (Grosz et al., 2020). This view occurred, in part, because of psychology's experimental roots in its early defining era of behaviorism and is joined by the more modern issue of the inaccurate 1:1 conflating of explanatory and causal work accompanied by the (still ever-present) misconception that causality can only be established via experimental manipulations. Accordingly, stereotypical explanatory work is generally seen less frequently in personality development research due to the obvious ethical and practical restraints concerning the manipulation of one's life circumstances and personality; phenomena of interest often not having unidirectional, tractable causes (Pearl & Mackenzie, 2018; Yarkoni, 2020); and the difficulty of attributing a nearly endless number of possible causes/explanations to the behavior of specific individuals (Mõttus et al., 2020). Instead, it is often the case that explanations resemble relatively more detailed and organized descriptions (Yarkoni, 2020), which can make the distinction between the different forms of research somewhat more ambiguous (Mõttus et al., 2020). However, explanatory research remains as crucial as description and prediction – it just often takes a form that is different than those with more conventional views would like, but, importantly, this makes it equally as likely to benefit from the inclusion of individual differences in sigma.

For instance, levels and changes in traits are the primary variables used to understand and explain how personality is associated with outcomes, interacts with one's environment, and is a consequential force in people's lives. It is possible to do this because *individual differences* in these metrics are quantified. For example, if the average level of conscientiousness in a sample is used to predict health status, and individuals range from very low to very high ratings of health status, then conscientiousness will not be associated with health. Rather, it is the individual associations of people's own levels of conscientiousness and health status, and the variability within these measures, that allow meaningful effects to be quantified. Then, underlying pathways for these effects can be investigated, and these investigations have indeed sometimes been fruitful in identifying mediating mechanisms (e.g., health behaviors or inflammation; Wright et al., 2022). Though, these effects are often small and explain minimal variance. The reason for this is almost certainly multidetermined: psychological processes are complex and isolating the influence of a few variables will undoubtedly only explain a small portion of variance (Mõttus et al., 2020); the investigated pathways apply to some, but not all, individuals; and other important explanatory variables are not being included. As such, this suggests other individual difference metrics can similarly be leveraged and the collective amount of variance explained could be meaningfully greater than when using any single metric alone. In this study, the degree of individual differences in sigma was as notable as that of initial trait levels and total changes in traits across time. Ignoring this valuable source of variance should be expected to have similar consequences that overlooking variance due to trait levels or changes in traits would have – which, at least in personality research abiding by a trait perspective, would mean discovered pathways shown to be important would have instead remained undetected.

Additionally, instead of only examining the isolated effects of multiple individual differences in personality and viewing the collective amount of variance they explain, it is crucial to investigate how their associations may be contingent upon one another. For example, although the conscientiousness-health link is among the more robust and frequently emerging associations in research, this effect is not always found (e.g., Wright & Jackson, 2023a). Individuals with high levels of conscientiousness may have poor health, and individuals with low levels may have excellent health. This not only highlights that a mechanism believed to explain an association or causally relate two variables need not be deterministic, but further that there are other factors that have instead led these individuals to have the health they do. Importantly, these do not need to be personality factors; however, someone does not just have the level of health that they have. Rather, they engage in certain health behaviors such as frequent exercise or not drinking, regularly eat a particular diet of foods, potentially have some pre-existing medical conditions, and/or were born with certain genetic predispositions that all likely influence their later health to some degree (Friedman, 2019; Graham et al., 2018; Kotov et al., 2010; Lunn et al., 2014; Terracciano et al., 2017; Turiano, Whiteman, et al., 2012; Wright et al., 2022). Given the many plausible mechanisms by which one's health can be affected, and the associations that personality typically has with the components in these pathways, it is sensible that personality can indirectly impact health via its role in one or more of these pathways.

Although these paths were not investigated in the present study, how the associations of multiple individual differences in personality with health status may differentially emerge based on other aspects of one's personality were examined. It was found that having lower-than-average increases in conscientiousness, or rather decreasing in this trait, was associated with below-average health status. Furthermore, interactions with sigma indicated that having greater-than-average within-person variability exacerbated this association and predicted even lower ratings of health status. Relatedly, having lower-than-average within-person variability nullified the negative association of slopes and predicted the highest ratings of health among people with decreases in conscientiousness. Then, and although this was only observed once for conscientiousness, but was frequently found for the other traits, greater-than-average sigma values would conversely strengthen positive associations that changes in traits had with health status, thus predicting the highest ratings, whereas having less within-person variability would predict slightly lower levels of health. This pattern of effects paints a more complex story than the simple one posited in the introduction. For individuals with attributes that are, on average, positive and beneficial for predicting an outcome, having greater within-person variability is similarly beneficial. Then, the opposite is true for those with attributes negatively associated with an outcome: greater variability is associated with even worse levels of the outcome.

This suggests the ways this within-person variability manifests for people may differ based on their other personality characteristics. Furthermore, considering this individual difference was correlated across all traits, this suggests this is a property of the person rather than a trait-specific manifestation. For instance, for people with trait levels and changes in traits positively associated with an outcome, their greater degree of within-person variability may serve to increase their adaptability in their environment. Therefore, even in the face of undesirable circumstances or external factors, they can leverage this quality and be relatively unaffected, or benefit even more, by being able to adjust to the new environmental demands instead of being rigidly inflexible and unable to adapt. In comparison, for people with personality characteristics that are negatively associated with an outcome, their greater variability may rather reflect a general lack of stability in their sense of self or behavioral routines. This inconstancy is thus not necessarily a tool they can use to navigate through their environment, as it depends on the nature of the personality characteristics from which it varies around, which are not beneficial in this instance. Overall, by isolating these associations, mapping them onto personality metrics already established in the literature, identifying the circumstances under which they arise, and examining their potential boundary conditions, the nature of how personality influences one's life can be better understood. Even with only one outcome, the possibilities for how sigma may influence how someone's personality is associated with an outcome and the nuances in those pathways are numerous. It seems likely that this would be the rule rather than the exception for other research in personality development.

4.2.5 Summary

Personality psychology is at the forefront of the intersection of advanced analytical approaches, rich sources of data, intricate study designs, and impactful research questions. Unfortunately, the utilization of these factors is imbalanced across the various subfields and even across research within a single subfield that is motivated by one goal versus another goal. This has led to disproportional progress in some areas; a lack of empirical work with robust, overarching theoretical frameworks; a lag in incorporating novel methods; and difficulty in obtaining a cumulative body of research. A potential solution is to leverage a new individual difference variable whose impact transcends boundaries of empirical versus theoretical work and has clear implications for research motivated by a goal of description, prediction, explanation, or some combination of the three.

Overall, ignoring an individual difference with implications such as helping researchers better understand the complexity of personality, more accurately portraying how it is associated with outcomes, reliably uncovering how it changes in light of other constructs, and strengthening theoretical frameworks could have profound consequences. The unifying advancements made in descriptive, predictive, and explanatory research for personality dynamics from quantifying and incorporating individual differences in within-person variability into empirical studies and theoretical frameworks should be no less expected to happen in personality development research. Even as an initial first step, the present study has underscored the potential that individual differences in sigma have for solving pieces of the personality development puzzle.

4.3 No, Heterogeneous Longitudinal Within-Person Variance Cannot Continue to be Treated as Homogeneous

As recent as only mere decades ago, there was still active debate as to if and to what extent personality remained stable versus was capable of change across the adult lifespan (Baltes, 1987; Brim & Kagan, 1980; Caspi & Bem, 1990; Caspi & Roberts, 1999; Conley, 1984, 1985; Costa & McCrae, 1994; Helson & Kwan, 2000; McCrae & Costa, 1990; Roberts & Chapman, 2001; Robins et al., 2001; Spiro et al., 2000). Early research indicated personality is mostly stable across time, both in terms of rank-order stability and mean-level changes (Conley, 1984, 1985; Costa & McCrae, 1994; Finn, 1986; Roberts & DelVecchio, 2000; Robins et al., 2001; Spiro et al., 2000). Continued research, motivated by theoretical and empirical work highlighting that individual differences in internal (e.g., genetics; Caspi & Roberts, 2001; Levenson & Crumpler, 1996) and external (e.g., life events; Neyer & Asendorpf, 2001; Roberts et al., 2003) factors likely foster average and unique changes in personality, and aided by methodological advancements that allowed for the modeling of interindividual differences in intraindividual change (McArdle & Epstein, 1987; Meredith & Tisak, 1990; Rogosa et al., 1982; von Eye & Nesselroade, 1992), then established there are indeed mean-level personality changes as well as individual differences in these changes (e.g., Mroczek & Spiro, 2003). Since then, it has become commonplace to allow interindividual variability in personality change across time, incorporate these individual differences into theories (e.g., Roberts & Mroczek, 2008; Roberts et al., 2008; Wrzus & Roberts, 2017), explore their boundary conditions and trait-specific nuances (e.g., Schwaba & Bleidorn, 2018), investigate their underlying mechanisms (e.g., Denissen et al., 2019; Specht et al., 2011), and use them to predict important outcomes (e.g., Hoff et al., 2021; Hounkpatin et al., 2018; Mroczek & Spiro, 2007; Wright & Jackson, 2023a).

Now, imagine if after the discovery of individual differences in mean-level changes, future research decided that, although this variability is present and can be meaningfully quantified, it was not important to regularly include it in empirical studies and continue exploring its implications. It is not hyperbole to say that personality, and even psychology in general, would be unrecognizable as a field today. Of course, it is difficult to discern which new empirical discovery has the potential to advance knowledge beyond what is conceivable at the time and change the future landscape of a scientific field. This is admittedly more easily recognized in hindsight, and the lack of the comparable counterfactual of a situation in which a discovery was instead overlooked and not further pursued makes it difficult to quantify its expected impact. Thankfully, the similarities between individual differences in mean-level changes and within-person residual variability, in conjunction with existing evidence that highlights the importance of within-person variability, can be used to make the case that continuing to treat heterogeneous variability around personality trajectories as homogeneous and/or meaningless noise with no implications would be a dire mistake.

First, in the present study, it was found that the degree of between-person heterogeneity for within-person variability around trajectories of the Big Five traits was comparable to initial trait levels and total changes in traits across time. Ultimately, this serves to contextualize and benchmark this variability by comparing it to the two most commonly examined individual differences in personality development. If, by default, individual differences in intercepts and slopes are quantified, then not modeling another individual difference with similar a degree of heterogeneity not only ignores potentially meaningful variance, but it further harms the precision of inferences as this variance is instead attributed to error or adds noise to other parameter estimates. For instance, even not including random effects for commonplace parameters such as intercept, slope, and covariance terms reduces power, produces biased estimates of standard errors, and inflates Type I error rates (Barr et al., 2013; Bell et al., 2019; Oberauer, 2022). Similarly, past research has found that in the case of truly heterogeneous within-person variances, treating them as homogenous produces biased estimates for variance parameters and standard errors of fixed effects (Hamel et al., 2012; Jahng & Wood, 2017; Leckie et al., 2014). A MELSM will produce more efficient standard errors and additionally result in less overall shrinkage (Kapur et al., 2015; Williams & Mulder, 2020). In the current study, the models that allowed individuals to differ in their degree of within-person residual variability were, without exception, the better-fitting models compared to those that constrained this to be homogeneous.

Second, changes in personality are often modeled linearly or quadratically in personality development research (c.f. Bleidorn et al., 2022; Fraley & Roberts, 2005; Wright & Jackson, 2023b), and are done so for the entire sample. Nevertheless, research beyond normative mean-level development suggests this is inappropriate. For example, past work has found substantial individual differences in best-fitting model forms for trajectories of the Big Five traits (Wright & Jackson, 2023d). This highlights the degree of heterogeneous patterns of change and suggests there are distinctive, non-shared processes underlying these patterns. This not only challenges the descriptive and applied value of average trajectories obtained from imposing the constraint of an identical model form across all individuals, but even further calls into question the inherent validity of these trajectories that have no means of accounting for nor quantifying the degree to which people’s development is satisfactorily approximated by this single model form. However, the reality is personality will continue to be predominantly modeled with linear or quadratic model forms. Although it has been found these are not the model forms that best fit majority of individuals in a sample (Wright & Jackson, 2023d), this is less of an issue if people are allowed to vary in how well they adhere to these forms. By relaxing the homoscedasticity assumption, individuals can differ in how well their actual trajectory is approximated by their line of best fit, as indicated by their person-specific degree of residual variance. Thus, the benefits afforded by using linear or quadratic trajectories can still be enjoyed, but now individual differences in within-person variability around these trajectories can also be quantified. These can then be used as indicators of individual-level model fit, highlight heterogeneity in personality development, and improve the precision of other parameters by accounting for a source of variance.

Lastly, and although it is difficult to accurately quantify or foreshadow its true impact, there would almost certainly be a significant loss of potential knowledge and slowed progress in understanding the complexity of personality development. While hypothetical parallels with individual differences in mean-level changes and their impact on personality development research can be drawn, a more direct comparison can be made with individual differences in within-person variability for personality dynamics research. As previously noted, this field flourished with the help of this newly validated construct. Importantly, all forms of research were necessary to conduct to holistically understand this within-person variability, and they all benefitted as well. As a result of the new collection of descriptive, predictive, and explanatory research, a cycle whereby theories were created to incorporate this variability as an important individual difference and new empirical evidence was used to update existing and inspire new theories naturally took place (Berry & Jobe, 2002; Blum et al., 2018; Cervone, 2005; Hooker, 2002; Mischel & Shoda, 1995). New methods and analytical frameworks were likewise created and evolved to equip researchers with the tools they needed to address their increasingly complex questions and hypotheses (e.g., Blozis, 2022; Hedeker et al., 2008; Nestler, 2022; Ram et al., 2012). Ultimately, this type of innovation is crucial to scientific progress.

In sum, the ease of following convention in personality development research by treating within-person variability around trajectories as homogeneous is not worth the statistical imprecision, sacrifice of quantifying meaningful variance, continued gaps in theoretical frameworks, and inability to solve missing pieces of the personality puzzle. People differ in the amount of longitudinal within-person residual variability they have; not quantifying this harms statistical inference and actively impedes the advancement of scientific knowledge.

4.4 Limitations & Future Directions

Although this study made an important step in the right direction for examining the nature of individual differences in within-person variability for personality development, future work can improve upon it and continue to explore various investigations and applications of this individual difference. This can be done in multiple ways. First, the temporal (in)stability of a psychological construct, including personality, is assumed to be comprised of two processes – within-person variability and temporal dependency (Jahng et al., 2008). Empirically, within-person variability can be quantified using traditional metrics such as intraindividual standard deviations or with approaches such as MELSMs. For temporal dependency, this can be captured with an autocorrelation (Cowdry et al., 1991; Stein, 1996), which allows one to quantify the persistence of a construct across time. Typically, both processes are believed to be more consequential for short time periods, such as in data obtained via experience sampling methods (Jahng et al., 2008; Sosnowska et al., 2019; Voelkle & Wagner, 2017). However, as seen in the current study, convention is not necessarily the best guide for conducting research. Individual differences in autocorrelations have indeed already been shown to exist for personality constructs in intensive longitudinal data (e.g., Jahng & Wood, 2017; Nestler, 2022; Vansteelandt & Verbeke, 2016). Furthermore, models that constrain the autocorrelation to be equal across people will have biased parameter estimates if they are truly heterogeneous (Hamel et al., 2012; Jahng & Wood, 2017; Ma et al., 2020). Thus, quantifying the degree to which there are individual differences in autocorrelations for longitudinal personality development, in conjunction with individual differences in sigma, could similarly prove to be a fruitful endeavor.

Second, the current study used manifest, or observed, variables. Although MELSMs are rarely ever fit with latent variables (e.g., Geukes et al., 2017; Goldstein et al., 2018; Hedeker et al., 2008, 2012; Kapur et al., 2015; Li & Hedeker, 2012; Nestler, 2022; Rast et al., 2012; Rast & Ferrer, 2018; Williams et al., 2019, 2020), a latent variable MELSM has been developed (Blozis, 2022). When scores include measurement error, sigma is a combination of the within-person residual variance and error variance. This can impact model fit and inflate the magnitude of the person-level residual variance (Blozis, 2022). However, if it can be assumed that measurement error is equally present for each individual, then the magnitude of the variability around the sample-level sigma value is unaffected, as standard deviation is invariant to change in origin (i.e., addition/subtraction of a constant). Notably, findings from a study that compared estimates from manifest and latent variable MELSMs (Blozis, 2022) might suggest this assumption is not so far-fetched. In this study, it was certainly true that the manifest variable MELSMs had larger sample-level sigma values than the latent variable MELSMs (Blozis, 2022). However, an interesting pattern emerges when examining the model-provided estimates for both models. In the latent variable MELSMs, the between-person variability around the sample-level sigma estimates is actually slightly larger than it is in the manifest variable MELSMs (parameter ϕ_a in

Tables 1-2). Additionally, the standard errors around both the fixed and random effects for sigma are larger in the latent variable MELSMs (the values in the parentheticals immediately following the parameter ϕ_a in Tables 1-2). Thus, when this within-person residual variability is of interest, conclusions from latent variable MELSMs are likely comparable to manifest variable MELSMs, with person-level heterogeneity possibly being larger in the former and some estimates from the latter perhaps being, somewhat paradoxically, more precise.

Third, only linear trajectories were fit in the current study. Many individuals are not best fit by a linear slope and their development is rather better approximated by a cubic or nonlinear trajectory (Wright & Jackson, 2023d). For these individuals, their sigma values will be larger than those for people that adhere quite well to a linear trajectory. Although this is one of the benefits of MELSMs, such that these individuals can potentially be identified, it would be worthwhile to fit models with additional forms (e.g., quadratic, cubic) to examine if these individuals maintain their relatively larger degree of residual variability or if their amount of residual variability decreases and others now have a greater amount. The former would suggest the reason for this large within-person variability is perhaps not due to underlying differences in types of trajectories, whereas the latter would suggest that individual differences in within-person residual variability reflect individual differences in model forms.

Fourth, since I opted to maximize the amount of data in each dataset, participants were included regardless of the number of waves they had. As previously noted, the use of MELSMs to examine within-person variability has typically only been done in studies with intensive longitudinal data. This type of data is usually characterized by a large number of repeated assessments per individual and a small to modest number of participants. The opposite is often true of longitudinal panel data: the number of participants often substantially outnumbers their number of waves. Simulations have shown that having a greater number of repeated assessments is more consequential than having more subjects in terms of power to detect intraindividual effects (Rast & Ferrer, 2018; Walters et al., 2018). The estimation of fixed effects is relatively unaffected by the joint Level 1 to Level 2 sample sizes, although having less data increases uncertainty for random effects (Williams et al., 2019). However, increasing the number of subjects does also increase power (Walters et al., 2018), and as model complexity increases, the requirements for the number of participants likely increase more quickly than the requirements for number of repeated measures (Rast & Ferrer, 2018). Furthermore, past work has successfully used MELSMs in data containing a few hundred subjects and five repeated measures per subject (Kapur et al., 2015; Williams et al., 2019). Thus, given that the sample sizes of our datasets ranged from 8,303 to 55,584, this could help mitigate some of the issues surrounding having few measurement occasions. To confirm this, simulation studies could examine the effects of varying the number of subjects and measurement occasions for MELSMs in longitudinal panel data, similar to what has been done for intensive longitudinal data. Additionally, future research would benefit regardless from having more measurement occasions per person, as this will lead to more precise estimates and help validate the findings in the current study.

Fifth, variance estimates of bounded variables are a function of the location of the mean (Baird et al., 2006; Eid & Diener, 1999; Kalmijn & Veenhoven, 2005; Rouder et al., 2008). This is not inherently an issue, but when bounded variables are the constructs of interest in a MELSM – such as personality traits that are nearly always constrained in their range of possible values – it may have implications for correlations among the random effects (Rast & Ferrer, 2018). These correlations may reflect a substantive effect such that variability in individuals' location and scale values for some variable(s) are meaningfully related to one another. Conversely, they may instead be a design artifact and rather reflect the constraint inherent in variance estimates that is dependent upon their mean (Rast & Ferrer, 2018; Williams et al., 2019). For instance, large positive or negative correlations between the location and scale parameters likely reflect floor or ceiling effects, respectively, and not a theoretically meaningful relationship. Of note, though, this can still be informative for future research, which further underscores the value of quantifying these effects with statistical frameworks such as MELSMs. With a dramatic enough ceiling/floor effect, this should hypothetically result in a smaller magnitude of both the sample-level residual variance estimate and the individual differences around this residual variability. For example, if a ceiling effect is found for a trait, such as was the case for conscientiousness in multiple datasets, then people are inherently restricted in their variability of scores across time. If most of a sample has their datapoints concentrated in some common vector space on the X-Y plane near the upper (or lower) boundary of the range of Y values, it is likely they are all well fit by a linear trajectory and their datapoints across time adhere quite closely to that trajectory. Accordingly, the magnitude of the between-person variability for the person-level residual variances will also be smaller in magnitude. However, it is unclear how the coefficient of variation would compare to a value obtained from data without a ceiling (or floor) effect. Similar to the necessity of exploring latent variable MELSMs, future research should examine the implications that various measurement properties have on the within-person variance parameters in MELSMs.

Lastly, for the regression models predicting health status, it was typically the case that an individual's outcome measure co-occurred with their last personality measure. Considering health status is regularly administered in the survey for each dataset, whereas Big Five traits are sometimes assessed more infrequently, distal measures of health status were not needed for most participants. Therefore, the outcome was not completely separated in time from the person-level sigma and slope values, as these parameters are based on all waves of data for a participant. Using wholly distal outcome data would have been the most conservative test of sigma's predictive utility. However, the purpose of these analyses was to serve as an initial examination of potential associations person-level sigma values would have with an outcome. Now that these associations have been found with a more liberal test, future research can examine if distal associations between individual differences in sigma and health status likewise emerge. Studies with outcomes other than health status may benefit from first conducting a similar set of simpler tests before advancing to those that are more conservative, though. First identifying if any basic associations exist prior to testing for temporal associations and causal pathways ensures foundational descriptive knowledge that can guide future research is obtained. This prevents the waste of resources spent trying to quantify a meaningful association, identify the perfect combination of moderating/mediating mechanisms that significantly contribute to a model, and/or outline boundary conditions for an effect that just simply may not exist. Description is at the core of all research; it is essential that new studies do not skip this crucial stage.

4.5 Conclusion

In this study, I identified meaningful individual differences in within-person residual variability for trajectories of the Big Five traits across five large-scale, longitudinal datasets. The magnitude of the variability in this individual difference was comparable to that of initial trait levels and change in traits across time. Moreover, the multitude of associations that within-person residual variability had with other individual differences for personality traits, a widespread set of variables with established empirical and theoretical importance in the field, and an outcome that has some of the most robust associations with the Big Five traits leaves little room for uncertainty regarding its value in personality development research. Ultimately, this source of within-person variability in longitudinal personality research cannot continue to be treated as if it were a fixed quantity across individuals that has no substantive meaning. Through regularly incorporating this individual difference into new empirical work, statistical inferences will improve, novel findings will be discovered, and the core goals of personality development as a scientific field will flourish with new research designed to complement one another and seamlessly harmonize together.

References

Allemand, M., Zimprich, D., & Hertzog, C. (2007). Cross-sectional age differences and longitudinal age changes of personality in middle adulthood and old age. *Journal of personality*, *75*(2), 323-358.

Allport, G. W. (1937). *Personality: A Psychological Interpretation*. New York: Holt, Rinehart, & Winston.

Andersen, H. K. (2022). Equivalent approaches to dealing with unobserved heterogeneity in cross-lagged panel models? Investigating the benefits and drawbacks of the latent curve model with structured residuals and the random intercept cross-lagged panel model. *Psychological Methods, 27*(5), 730–751. https://doi.org/10.1037/met0000285

Asendorpf, J. B., & van Aken, M. A. (1991). Correlates of the temporal consistency of personality patterns in childhood. *Journal of Personality*, *59*(4), 689-703.

Asparouhov, T., & Muthén, B. (2022). Residual structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-31.

Atherton, O. E., Robins, R. W., Rentfrow, P. J., & Lamb, M. E. (2014). Personality correlates of risky health outcomes: Findings from a large Internet study. *Journal of Research in Personality*, *50*, 56-60.

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: reliability, validity, and associations with well-being. *Journal of personality and social psychology*, *90*(3), 512.

Baltes, P. B. (1987). Theoretical propositions of lifespan developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, 23, 611–626.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, 50, 471–507.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., ... & Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*(5), 503-528.

Beck, E. D., Condon, D., & Jackson, J. (2022). Interindividual age differences in personality structure. *European Journal of Personality*, https://doi.org/10.1177/08902070221084862

Beck, E. D., & Jackson, J. J. (2020). Consistency and change in idiographic personality: A longitudinal ESM network study. *Journal of Personality and Social Psychology*, *118*(5), 1080.

Beck, E. D., & Jackson, J. J. (2021). Within-person variability. In *The handbook of personality dynamics and processes* (pp. 75-100). Academic Press.

Beck, E. D., & Jackson, J. J. (2022). A mega-analysis of personality prediction: Robustness and boundary conditions. *Journal of Personality and Social Psychology, 122*(3), 523-553.

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & quantity*, *53*, 1051-1074.

Berry, J. M., & Jobe, J. B. (2002). At the intersection of personality and adult development. *Journal of Research in Personality, 36*(4), 283-286.

Bleidorn, W., Hopwood, C. J., & Lucas, R. E. (2018). Life events and personality trait change. *Journal of personality*, *86*(1), 83-96.

Bleidorn, W., Kandler, C., Riemann, R., Spinath, F. M., & Angleitner, A. (2009). Patterns and sources of adult personality development: Growth curve analyses of the NEO PI-R scales in a longitudinal twin study. *Journal of Personality and Social Psychology*, *97*(1), 142–155. https://doi.org/10.1037/a0015434

Bleidorn, W., Klimstra, T. A., Denissen, J. J., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality maturation around the world: a cross-cultural examination of social-investment theory. *Psychological science*, *24*(12), 2530–2540.

Bleidorn, W., Schwaba, T., Zheng, A., Hopwood, C. J., Sosa, S. S., Roberts, B. W., & Briley, D. A. (2022). Personality Stability and Change: A Meta-Analysis of Longitudinal Studies. *Psychological bulletin*. https://doi.org/10.1037/bul0000365

Blozis, S. A. (2022). A Latent Variable Mixed-Effects Location Scale Model with an Application to Daily Diary Data. *Psychometrika*, *87*(4), 1548-1570.

Blum, G. S., Rauthmann, J. F., Göllner, R., Lischetzke, T., & Schmitt, M. (2018). The nonlinear interaction of person and situation (NIPS) model: Theory and empirical evidence. *European Journal of Personality*, *32*(3), 286-305.

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

Bollen, K. A., & Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) Models A Synthesis of Two Traditions. *Sociological Methods & Research*, *32*(3), 336–383.

Bollen, K. A., & Zimmer, C. (2010). An overview of the autoregressive latent trajectory (ALT) model. *Longitudinal research with latent variables*, 153-176.

Brim, O. G., Jr., Kagan, J. (1980). Constancy and change: A view of the issues. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 1–25).

Cambridge, MA: Harvard University Press.

Bureau of Labor Statistics, U.S. Department of Labor, and National Institute for Child Health and Human Development. Children of the NLSY79, 1979-2016. (2020). *Produced and distributed by the center for human resource research (CHRR)*. Columbus, OH: The Ohio State University.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Inference: A Practical Information-Theoretic Approach*. 2nd Edition, Springer-Verlag, New York.

Caspi, A. (1998). Personality development across the life course. In W. Damon & N. Eisenberg (Eds.), Handbook of child psychology (Vol. 3, pp. 311–388). New York: Wiley.

Caspi, A., Bem, D. J. (1990). Personality continuity and change across the life course. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (1st ed., pp. 549–569). New York: Guilford Press.

Caspi, A., Roberts, B. W. (1999). Personality development across the life course. In L. A. Pervin and O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 300–326). New York: Guilford Press.

Caspi, A., Roberts, B. W. (2001). Personality development across the life: The argument for change and continuity. *Psychological Inquiry*, 12, 49-66.

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology*, *38*(4), 476.

Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, *12*, 267-288.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annu. Rev. Psychol.*, *56*, 423-452.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, *14*(3), 464-504.

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate behavioral research*, *34*(3), 315-346.

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. *Personality and Individual Differences*, *5*, 11-26.

Conley, J. J. (1985). Longitudinal stability of personality traits: A multitrait-multimethod-multioccasion analysis. *Journal of Personality and Social Psychology*, *49*, 1266-1282.

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and personality psychology compass*, *3*(3), 292-313.

Converse, P. D., Beverage, M. S., Vaghef, K., & Moore, L. S. (2018). Self-control over time: Implications for work, relationship, and well-being outcomes. *Journal of Research in Personality, 73,* 82–92. https://doi.org/10.1016/j.jrp.2017.11.002

Costa, P. T., McCrae, R. R. (1994). Set like plaster? Evidence for the stability of adult personality. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change* (pp. 21–40)? Washington, DC: American Psychological Association.

Cowdry, R. W., Gardner, D. L., O'Leary, K. M., Leibenluft, E., & Rubinow, D. R. (1991). Mood variability: a study of four groups. *The American journal of psychiatry*.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50*(6), 456–473. https://doi.org/10.1037/h0057173

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*(5), 879–894. https://doi.org/10.1037/a0035297

De Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology*, 91, 538–552.

Denissen, J. J. A., Luhmann, M., Chung, J. M., & Bleidorn, W. (2019). Transactions between life events and personality traits across the adult lifespan. *Journal of Personality and Social Psychology, 116*(4), 612–633. https://doi.org/10.1037/pspp0000196

Donnellan, M. B., Conger, R. D., & Burzette, R. G. (2007). Personality development from late adolescence to young adulthood: Differential stability, normative maturity, and evidence for the maturity–stability hypothesis. *Journal of Personality*, 75, 237–264.

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76*(4), 662-676.

Eysenck, H. J., & Eysenck, S. B. (2013). *Personality structure and measurement (psychology revivals)*. Routledge.

Finn, S. E. (1986). Stability of personality self-ratings over 30 years: Evidence for an age/cohort interaction. *Journal of Personality and Social Psychology*, *50*, 813-818.

Fleeson, W. (2001). Towards a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011-

1027.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13*(2), 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of personality*, *75*(4), 825-862.

Friedman, H. S. (2019). Neuroticism and health as individuals age. *Personality Disorders: Theory, Research, and Treatment*, *10*(1), 25-32. https://doi.org/10.1037/per0000274

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: a dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological review*, *112*(1), 60-74.

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C., & Back, M. D. (2017). Trait personality and state variability: Predicting individual differences in within-and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, *69*, 124-138.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.

Goldstein, H., Leckie, G., Charlton, C., Tilling, K., & Browne, W. J. (2018). Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. *Statistical methods in medical research*, *27*(11), 3478-3491.

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, *17*(1), 205-215.

Graham, E. K., Bastarache, E. D., Milad, E., Turiano, N. A., Cotter, K. A., & Mroczek, D. K. (2018). Physical activity mediates the association between personality and biomarkers of inflammation. *SAGE Open Medicine*, *6*. https://doi.org/10.1177/2050312118774990

Graham, E. K., Weston, S. J., Gerstorf, D., Yoneda, T. B., Booth, T. O. M., Beam, C. R., ... & Mroczek, D. K. (2020). Trajectories of big five personality traits: A coordinated analysis of 16 longitudinal samples. *European Journal of Personality*, *34*(3), 301-321.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243-1255.

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, *33*(3), 404–416. https://doi.org/10.1177/0049124104270220

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. https://doi.org/10.1037/a0038889

Hamaker, E. L., Mulder, J. D., & van IJzendoorn, M. H. (2020). Description, prediction and causation: Methodological challenges of studying child and adolescent development. *Developmental cognitive neuroscience*, *46*, 100867.

Hamel, S., Yoccoz, N. G., & Gaillard, J. M. (2012). Statistical evaluation of parameters estimating autocorrelation and individual heterogeneity in longitudinal studies. *Methods in Ecology and Evolution*, *3*(4), 731-742.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2006). Forty years on: teachers' assessments of children's personality traits predict self-reported health behaviors and outcomes at midlife. *Health psychology*, *25*(1), 57.

Hardy, J., & Segerstrom, S. C. (2017). Intra-individual variability and psychological flexibility: Affect and health in a National US sample. *Journal of Research in Personality*, *69*, 13-21.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627-634.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in medicine*, *31*(27), 3328-3336.

Heller, D., Komar, J., & Lee, W. B. (2007). The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin*, *33*(6), 898-910.

Helson, R., Jones, C., & Kwan, V. S. Y. (2002). Personality change over 40 years of adulthood: Hierarchical linear modeling analyses of two longitudinal samples. *Journal of Personality and Social Psychology*, 83, 752–766.

Helson, R., Kwan, V. S. Y. (2000). Personality change in adulthood: The broad picture and processes in one longitudinal study. In S. Hampson (Ed.), *Advances in personality psychology, Vol. 1* (pp. 77–106). Hove, England: Psychology Press.

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: a classification of data science tasks. *Chance*, *32*(1), 42-49.

Hill, P. L., Payne, B. R., Jackson, J. J., Stine-Morrow, E. A., & Roberts, B. W. (2014). Perceived social support predicts increased conscientiousness during older adulthood. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *69*(4), 543-547.

Hoff, K. A., Einarsdóttir, S., Chu, C., Briley, D. A., & Rounds, J. (2021). Personality Changes

Predict Early Career Outcomes: Discovery and Replication in 12-Year Longitudinal Studies. *Psychological Science*, *32*(1), 64-79. https://doi.org/10.1177/0956797620957998

Hollenbach, F. M., & Montgomery, J. M. (2020). Bayesian model selection, model comparison, and model averaging. *SAGE Handbook of Research Methods in Political Science and International Relations*, 937-960.

Hooker, K. (2002). New directions for research in personality and aging: A comprehensive model for linking levels, structures, and processes. *Journal of Research in Personality*, *36*(4), 318-334.

Hounkpatin, H. O., Boyce, C. J., Dunn, G., & Wood, A. M. (2018). Modeling bivariate change in individual differences: Prospective associations between personality and life satisfaction. *Journal of Personality and Social Psychology, 115*(6), 12-29.

Howell, R. T., Ksendzova, M., Nestingen, E., Yerahian, C., & Iyer, R. (2017). Your personality on a good day: How trait and state personality predict daily well-being. *Journal of Research in Personality*, *69*, 250-263.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hudson, N. W., & Roberts, B. W. (2016). Social investment in work reliably predicts change in conscientiousness and agreeableness: A direct replication and extension of Hudson, Roberts, and Lodi-Smith (2012). *Journal of Research in Personality*, *60*, 12-23.

Ibáñez, M. I., Viruela, A. M., Mezquita, L., Moya, J., Villa, H., Camacho, L., & Ortet, G. (2016). An investigation of five types of personality trait continuity: A two-wave longitudinal study of Spanish adolescents from age 12 to age 15. *Frontiers in psychology*, *7*, 512.

Jackson, J. J., & Beck, E. D. (2021). Personality development beyond the mean: Do life events shape personality variability, structure, and ipsative continuity?. *The Journals of Gerontology: Series B*, *76*(1), 20-30.

Jackson, J. J., Connolly, J. J., Garrison, S. M., Leveille, M. M., & Connolly, S. L. (2015). Your friends know how long you will live: A 75-year study of peer-rated personality traits. *Psychological Science*, *26*(3), 335-340.

Jackson, J. J., Thoemmes, F., Jonkmann, K., Lüdtke, O., & Trautwein, U. (2012). Military training and personality trait development: Does the military make the man, or does the man make the military?. *Psychological science*, *23*(3), 270-277.

Jahng, S., & Wood, P. K. (2017). Multilevel models for intensive longitudinal data with heterogeneous autoregressive errors: the effect of misspecification and correction with Cholesky transformation. *Frontiers in psychology*, *8*, 262.

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological methods*, *13*(4), 354.

Johnson, A. A., Ott, M. Q., & Dogucu, M. (2022). *Bayes rules!: an introduction to applied Bayesian modeling*. CRC Press.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: *Useful tools for structural equation modeling*. R package version 0.5-6. https://CRAN.R-project.org/package=semTools

Juster, F. T., & Suzman, R. (1995). An overview of the Health and Retirement Study. *The Journal of Human Resources*, 30, S7–S56. https://doi.org/10.2307/146277

Kalmijn, W., & Veenhoven, R. (2005). Measuring inequality of happiness in nations: In search for proper statistics. *Journal of Happiness Studies*, *6*, 357-396.

Kapur, K., Li, X., Blood, E. A., & Hedeker, D. (2015). Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: an application to ecological momentary assessment data. *Statistics in Medicine*, *34*(4), 630-651.

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multi-wave data. *Journal of Consulting and Clinical Psychology*, *63*, 52-59.

Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 243–263). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10409-008

Klimstra, T. A., Luyckx, K., Hale III, W. W., Goossens, L., & Meeus, W. H. (2010). Longitudinal associations between personality profile stability and adjustment in college students: Distinguishing among overall stability, distinctive stability, and within-time normativeness. *Journal of personality*, *78*(4), 1163-1184.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*(5), 768–821. https://doi.org/10.1037/a0020327

Kuper, N., Modersitzki, N., Phan, L. V., & Rauthmann, J. F. (2021). The dynamics, processes, mechanisms, and functioning of personality: An overview of the field. *British Journal of Psychology*, *112*(1), 1-51.

Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance–covariance components in two-level models. *Journal of Educational and Behavioral*

Statistics, *39*(5), 307-332.

Lehmann, R., Denissen, J. J. A., Allemand, M., & Penke, L. (2013). Age and gender differences in motivational manifestations of the Big Five from age 16 to 60. *Developmental Psychology, 49*(2), 365–383. https://doi.org/10.1037/a0028277

Lehnart, J., Neyer, F. J., & Eccles, J. (2010). Long-term effects of social investment: The case of partnering in young adulthood. *Journal of personality*, *78*(2), 639-670.

Letzring, T. D., Edmonds, G. W., & Hampson, S. E. (2014). Personality change at mid-life is associated with changes in self-rated health: Evidence from the Hawaii Personality and Health Cohort. *Personality and Individual Differences*, *58*, 60-64.

Levenson, M. R., & Crumpler, C. A. (1996). Three models of adult development. *Human Development, 39*(3), 135–149. https://doi.org/10.1159/000278429

Li, X., & Hedeker, D. (2012). A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine*, *31*(26), 3192-3210.

Lievens, F., Lang, J. W., De Fruyt, F., Corstjens, J., Van de Vijver, M., & Bledow, R. (2018). The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment. *Journal of Applied Psychology*, *103*(7).

Lodi-Smith, J., & Roberts, B. W. (2007). Social investment and personality: A meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism. *Personality and social psychology review*, *11*(1), 68-86.

Lönnqvist, J. E., Mäkinen, S., Paunonen, S. V., Henriksson, M., & Verkasalo, M. (2008). Psychosocial functioning in young men predicts their personality stability over 15 years. *Journal of Research in Personality*, *42*(3), 599-621.

Lucas, R. E. (2023). Why the Cross-Lagged Panel Model Is Almost Never the Right Choice. *Advances in Methods and Practices in Psychological Science*, *6*(1), 1-22.

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology*, 101, 847-861.

Lüdtke, O., Roberts, B. W., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: life paths, life events, and personality trait change at the transition to university life. *Journal of personality and social psychology*, *101*(3), 620-637.

Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-20.

Lüdtke, O., Robitzsch, A., & Wagner, J. (2018). More stable estimation of the STARTS model: A Bayesian approach using Markov chain Monte Carlo techniques. *Psychological Methods, 23*(3), 570–593. https://doi.org/10.1037/met0000155

Lunn, T. E., Nowson, C. A., Worsley, A., & Torres, S. J. (2014). Does personality affect dietary intake?. *Nutrition*, *30*(4), 403-409.

Ma, Q., Mermelstein, R., & Hedeker, D. (2020). A three-level mixed model to account for the correlation at both the between-day and the within-day level for ecological momentary assessments. *Health Services and Outcomes Research Methodology*, *20*, 247-264.

Magee, C., Buchtel, E. E., Human, L. J., Murray, D. R., & Biesanz, J. C. (2018). Is personality variability associated with adjustment?. *Journal of Research in Personality*, *72*, 22-43.

Magee, C. A., Heaven, P. C., & Miller, L. M. (2013). Personality Change Predicts Self-Reported Mental and Physical Health. *Journal of Personality*, *81*(3), 324-334.

McArdle, J. J., & Epstein, D. (1987). Latent Growth Curves within Developmental Structural Equation Models. *Child Development*, *58*(1), 110–133. https://doi.org/10.2307/1130295

McCabe, K. (2017). Within-Person Variability of Personality and Individual Differences. In: Zeigler-Hill, V., Shackelford, T. (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_1355-1

McCrae, R. R., Costa, P. T. (1990). *Personality in Adulthood*. New York: Guilford Press.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*, *102*(2), 246.

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.

Mroczek, D. K., & Spiro, A., 3rd (2003). Modeling intraindividual change in personality traits: findings from the normative aging study. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *58*(3), 153–165.

Mroczek, D. K., & Spiro, A. (2007). Personality change influences mortality in older men. *Psychological science*, *18*(5), 371-376.

Mueller, S., Wagner, J., Drewelies, J., Duezel, S., Eibich, P., Specht, J., Demuth, I., Steinhagen-Thiessen, E., Wagner, G. G., & Gerstorf, D. (2016). Personality development in old age relates to physical health and cognitive performance: Evidence from the Berlin Aging

Study II. *Journal of Research in Personality*, *65*, 94–108.

Murayama K., Gfrörer T. (2022). Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for model choice from causal inference perspective. PsyArXiv. https://doi.org/10.31234/osf.io/bt9xr

Murray, G., Allen, N. B., & Trinder, J. (2002). Longitudinal investigation of mood variability and the ffm: Neuroticism predicts variability in extended states of positive and negative affect. *Personality and Individual Differences*, *33*(8), 1217-1228.

Nesselroade, J. R. (2001). Intraindividual variability in development within and between individuals. *European Psychologist, 6*(3), 187–193.

Nestler, S. (2022). An extension of the mixed-effects growth model that considers between-person differences in the within-subject variance and the autocorrelation. *Statistics in Medicine*, *41*(3), 471-482.

Neyer, F. J., & Asendorpf, J. B. (2001). Personality relationship transaction in young adulthood. *Journal of Personality and Social Psychology*, *81*, 1190–1204.

Oberauer, K. (2022). The importance of random slopes in mixed models for Bayesian hypothesis testing. *Psychological Science*, *33*(4), 648-665.

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, *120*(4), 1013-1034.

Ozer, D. J., & Gjerde, P. F. (1989). Patterns of personality consistency and change from childhood through adolescence. *Journal of Personality*, 57, 483–507.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.

Penner, L. A., Shiffman, S., Paty, J. A., & Fritzsche, B. A. (1994). Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology, 66*(4), 712–721. https://doi.org/10.1037/0022-3514.66.4.712

Popper, K. R. (1959). *The logic of scientific discovery.* Basic Books.

Pusch, S., Mund, M., Hagemeyer, B., & Finn, C. (2019). Personality development in emerging and young adulthood: A study of age differences. European Journal of Personality, *33*(3), 245-263.

Quilty, L. C., Meusel, L. A. C., & Bagby, R. M. (2008). Neuroticism as a mediator of treatment response to SSRIs in major depressive disorder. *Journal of affective disorders*, *111*(1), 67-73.

Quine, W. V. O. (1980). *From a logical point of view: Nine logico-philosophical essays* (No. 566). Harvard University Press.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: http://www.R-project.org/

Ram, N., Conroy, D., Pincus, A., Hyde, A., & Molloy, L. (2012). Tethering theory to method: Using measures of intraindividual variability to operationalize individuals' dynamic characteristics. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 81–110). IAP Information Age Publishing.

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: tools for examining the development of dynamic characteristics and processes. *Psychology and aging*, *24*(4), 778–791. https://doi.org/10.1037/a0017915

Rast, P., & Ferrer, E. (2018). A mixed-effects location scale model for dyadic interactions.

Multivariate behavioral research, *53*(5), 756-775.

Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research*, *47*(2), 177-200.

Revelle, W. (2021). psych: *Procedures for psychological, psychometric, and personality research* (R package Version 2.1.3). Northwestern University. https://CRAN.R-project.org/package=psych

Roberts, B. W. (1997). Plaster or plasticity: Are adult work experiences associated with personality change in women? *Journal of Personality, 65*(2), 205-232.

Roberts, B. W., Caspi, A., & Moffitt, T. E. (2001). The kids are alright: Growth and stability in personality development from adolescence to adulthood. *Journal of personality and social psychology*, *81*(4), 670-683.

Roberts, B. W., Chapman, C. N. (2001). Change in dispositional well-being and its relations to role quality: A 30-year longitudinal study. *Journal of Research in Personality*, 34, 26-41.

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

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological science*, *2*(4), 313-345.

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current directions in psychological science*, *17*(1), 31-35.

Roberts, B. W. & Nickel, L. B. (2021). Personality development across the life course: A neo-socioanalytic perspective. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (pp. 259–283). The Guilford Press.

Roberts, B. W., Robins, R. W., Caspi, A., & Trzesniewski, K. H. (2003). Personality trait development in adulthood. In J. T. Mortimer & M. Shanahan (Eds.), Handbook of the life course (pp. 579–595). New York: Plenum.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological bulletin*, *132*(1), 1-25.

Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 375–398). The Guilford Press.

Robins, R. W., Caspi, A., & Moffitt, T. E. (2002). It's not just who you're with, it's who you are: Personality and relationship experiences across multiple relationships. *Journal of personality*, *70*(6), 925-964. https://doi.org/10.1111/1467-6494.05028

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, 69, 617–640.

Rogosa, D.R., Brandt, D., & Zimowski, M.F. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726-748.

Rouder, J. N., Tuerlinckx, F., Speckman, P., Lu, J., & Gomez, P. (2008). A hierarchical approach for fitting curves to response time measurements. *Psychonomic Bulletin & Review*, *15*(6), 1201-1208.

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., ... &

McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, *117*(15), 8398-8403.

Scherpenzeel, A. C., & Das, M. (2010). "True" longitudinal and probability-based internet panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet: Advances in applied methods and research strategies* (pp. 77–104). Taylor and Francis.

Schwaba, T., & Bleidorn, W. (2018). Individual differences in personality change across the adult life span. *Journal of personality*, *86*(3), 450-464.

Schwaba, T., Robins, R. W., Grijalva, E., & Bleidorn, W. (2019). Does openness to experience matter in love and work? Domain, facet, and developmental evidence from a 24-year longitudinal study. *Journal of Personality*, *87*(5), 1074-1092.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310.

Sivula, T., Magnusson, M., Matamoros, A. A., & Vehtari, A. (2020). Uncertainty in Bayesian leave-one-out cross-validation based model comparison. *arXiv preprint*.

Socio-Economic Panel. (2020). Socio-Economic Panel (SOEP) data for years 1984–2020 (v37).

Sosnowska, J., Kuppens, P., De Fruyt, F., & Hofmans, J. (2019). A dynamic systems approach to personality: The Personality Dynamics (PersDyn) model. *Personality and individual differences*, *144*, 11-18.

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.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology, 101*(4), 862-882.

Spiro, A., III, Butcher, J. N., Levenson, M. R., Aldwin, C. M., Bossé, R. (2000). Change and stability in personality: A 5-year study of the MMPI-2 in older men. In J. N. Butcher (Ed.), *Basic sources on the MMPI-2.* Minneapolis: University of Minnesota Press.

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

Stein, K. F. (1996). Affect instability in adults with a borderline personality disorder. *Archives of Psychiatric Nursing*, *10*(1), 32-40.

Takahashi, Y., Edmonds, G. W., Jackson, J. J., & Roberts, B. W. (2013). Longitudinal correlated changes in conscientiousness, preventative health-related behaviors, and self-perceived physical health. *Journal of Personality, 81*(4), 417-427.

Terracciano, A., Stephan, Y., Luchetti, M., & Sutin, A. R. (2018). Cognitive impairment, dementia, and personality stability among older adults. *Assessment*, *25*(3), 336-347.

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T. Jr. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging*, *20*, 493–506.

Terracciano, A., McCrae, R. R., & Costa Jr, P. T. (2010). Intra-individual change in personality stability and age. *Journal of research in personality*, *44*(1), 31-37.

Terracciano, A., Stephan, Y., Luchetti, M., Albanese, E., & Sutin, A. R. (2017). Personality traits and risk of cognitive impairment and dementia. *Journal of Psychiatric Research*, *89*, 22-27. https://doi.org/10.1016/j.jpsychires.2017.01.011

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: a study of cultural mechanisms of personality. *Journal of Personality and Social Psychology*, *85*(4), 723.

Turiano, N.A., Pitzer, L., Armour, C., Karlamangla, A., Ryff, C.D., & Mroczek, D.K. (2012). Personality trait level and change as predictors of health outcomes: Findings from a national study of Americans (MIDUS). *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *67*(1), 4-12.

Turiano, N. A., Whiteman, S. D., Hampson, S. E., Roberts, B. W., & Mroczek, D. K. (2012). Personality and substance use in midlife: Conscientiousness as a moderator and the effects of trait change. *Journal of Research in Personality, 46*(3), 295-305.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24,* 637–657.

van Aken, M. A. G., Denissen, J. J. A., Branje, S. J. T., Dubas, J. S., & Goossens, L. (2006). Midlife concerns and short-term personality change in middle adulthood. *European Journal of Personality*, *20*(6), 497–513. https://doi.org/10.1002/per.603

van Scheppingen, M. A., Jackson, J. J., Specht, J., Hutteman, R., Denissen, J. J. A., & Bleidorn, W. (2016). Personality trait development during the transition to parenthood: A test of social investment theory. *Social Psychological and Personality Science, 7,* 452– 462.

Vansteelandt, K., & Verbeke, G. (2016). A mixed model to disentangle variance and serial autocorrelation in affective instability using ecological momentary assessment data. *Multivariate Behavioral Research*, *51*(4), 446-465.

Vazire, S., & Sherman, R. A. (2017). Introduction to the special issue on within-person variability in personality. *Journal of Research in Personality*, *69*, 1-3.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, *27*, 1413-1432.

Vize, C. E., Sharpe, B. M., Miller, J. D., Lynam, D. R., & Soto, C. J. (2022). Do the Big Five personality traits interact to predict life outcomes? Systematically testing the prevalence, nature, and effect size of trait-by-trait moderation. *European Journal of Personality*. https://doi.org/10.1177/08902070221111857

Voelkle, M. C., & Wagner, J. (2017). Analyzing personality change: From average trajectories to within-person dynamics. In *Personality development across the lifespan* (pp. 497-516). Academic Press.

von Eye, A., & Nesselroade, J. R. (1992). Types of change: application of configural frequency analysis in repeated measurement designs. *Experimental aging research*, *18*(3-4), 169-183. https://doi.org/10.1080/03610739208260357

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*, 192-196.

Wagner, J., Becker, M., Lüdtke, O., & Trautwein, U. (2015). The first partnership experience and personality development: A propensity score matching study in young adulthood. *Social Psychological and Personality Science*, *6*(4), 455-463.

Walters, R. W., Hoffman, L., & Templin, J. (2018). The power to detect and predict individual differences in intra-individual variability using the mixed-effects location-scale model. *Multivariate behavioral research*, *53*(3), 360-374.

Watson, N., & Wooden, M. (2012). The HILDA Survey: A case study in the design and development of a successful Household Panel Study. *Longitudinal and Life Course Studies*, *3*(3), 369–381. https://doi.org/10.14301/llcs.v3i3.208

Williams, D. R., Martin, S. R., Liu, S., & Rast, P. (2020). Bayesian Multivariate Mixed-Effects Location Scale Modeling of Longitudinal Relations Among Affective Traits, States, and Physical Activity. *European Journal of Psychological Assessment*, *36*(6), 981–997.

Williams, D. R., & Mulder, J. (2020). Bayesian hypothesis testing for Gaussian graphical models: Conditional independence and order constraints. *Journal of Mathematical Psychology*, *99*, 102441.

Williams, D. R., Zimprich, D. R., & Rast, P. (2019). A Bayesian nonlinear mixed-effects location scale model for learning. *Behavior research methods*, *51*(5), 1968-1986.

Woodrow, H. (1932). Quotidian variability. *Psychological Review, 39*(3), 245-256.

Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment, 31*(4), 502–515. https://doi.org/10.1037/pas0000617

Wright, A. J., & Jackson, J. J. (2023b). Are some people more consistent? Examining the stability and underlying processes of personality profile consistency. *Journal of Personality and Social Psychology, 124*(6), 1314-1337. https://doi.org/10.1037/pspp0000429

Wright, A. J., & Jackson, J. J. (2022a). Childhood temperament and adulthood personality differentially predict life outcomes. *Scientific reports*, *12*(1), 1-9. https://doi.org/10.1038/s41598-022-14666-0

Wright, A. J., & Jackson, J. J. (2023a). Do changes in personality predict life outcomes? *Journal of Personality and Social Psychology.*

Wright, A. J., & Jackson, J. J. (2023d). Examining individual differences in best-fitting model forms for Big Five trajectories. [Manuscript in preparation]

Wright, A. J., & Jackson, J. J. (2023e). Initiation of drug and alcohol use and personality development during adolescence. *European Journal of Personality*, *34*(7), 375-401. https://doi.org/10.1177/08902070221090107

Wright, A. J., & Jackson, J. J. (2023f). Is parent personality associated with child outcomes? A response surface analysis approach. *Infant and Child Development*. https://doi.org/10.1002/icd.2395

Wright, A. J., Weston, S. J., Norton, S., Voss, M., Bogdan, R., Oltmanns, T. F., & Jackson, J. J. (2022). Prospective self- and informant-personality associations with inflammation, health behaviors, and health indicators. *Health Psychology, 41*(2), 121-133. https://doi.org/10.1037/hea0001162

Wright, A. J., & Jackson, J. J. (2023c). The associations between life events and person-centered personality consistency. *Journal of Personality*, 1-18. https://doi.org/10.1111/jopy.12802

Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. *Personality and Social Psychology Review, 21*(3), 253-277.

Yarkoni, T. (2020). Implicit realism impedes progress in psychology: Comment on Fried (2020). *Psychological Inquiry*, *31*(4), 326-333.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122.