Supplemental Materials

Using Within-Person Change in Three Large Panel Studies to Estimate Personality Age Trajectories

Table S1

Outline of the Supplemental Materials

Supplemental Materials for Study 1: Household, Income and Labour Dynamics in Australia (HILDA) Survey

Sample Description and Psychometric Analyses

Table S2

Sample Characteristics by Trait in the Household, Income and Labour Dynamics in Australia (HILDA) Survey

Trait	*n*	Age		% female
		M	*SD*	
Neuroticism	15,083	45.23	18.01	53.30
Extraversion	15,014	45.18	17.95	53.24
Conscientiousness	15,075	45.20	17.98	53.19
Agreeableness	15,168	45.19	18.03	53.36
Openness	15,009	45.13	17.95	53.21

Table S3

Wave-Specific Psychometric Properties in the Household, Income and Labour Dynamics in Australia (HILDA) Survey

Parameter	2005	2009	2013	2017
		Neuroticism		
n	9,075	10,350	13,336	12,663
α	.76	.77	.77	.77
ω	.81	.81	.82	.80
		Extraversion		
n	8,972	10,276	13,251	12,585
α	.67	.68	.69	.70
ω	.69	.70	.71	.72
		Conscientiousness		
n	9,049	10,327	13,326	12,643
α	.78	.78	.78	.78
ω	.80	.81	.80	.81
		Agreeableness		
n	9,131	10,428	13,425	12,760
α	.75	.75	.75	.76
ω	.83	.84	.81	.84
		Openness		
n	8,994	10,266	13,263	12,609
α	.70	.72	.71	.73
ω	.74	.75	.74	.75

Fixed Effects Modeling

Table S4

Parameter Estimates for the Effect of the First Wave on Personality in the Household, Income and Labour Dynamics in Australia (HILDA) Survey

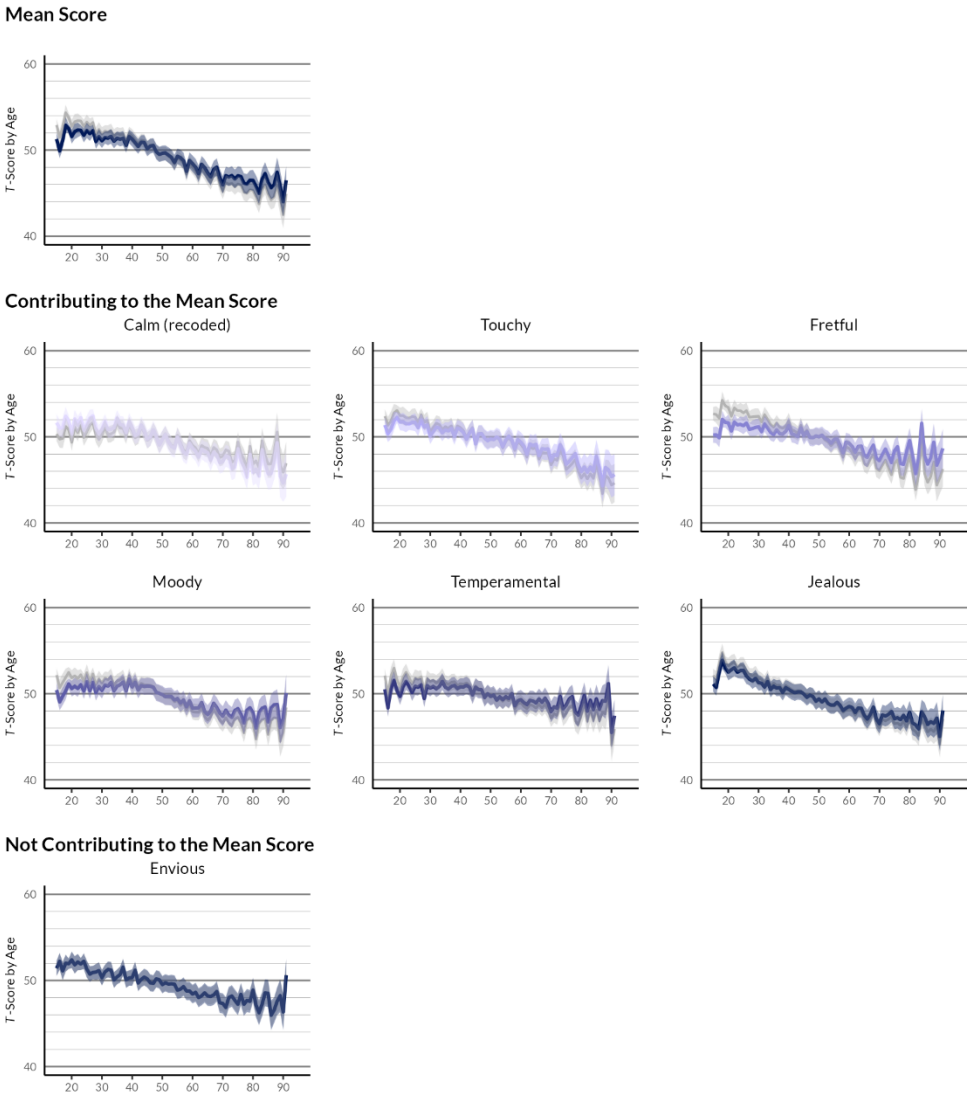
Model	Estimate	*SE*	*t*	*p*
	Neuroticism			
Mean score	0.441	0.102	4.327	< .001
Contributing to the mean score				
Calm (recoded)	−0.361	0.129	−2.796	.005
Touchy	0.279	0.134	2.076	.038
Fretful	0.670	0.125	5.366	< .001
Moody	0.468	0.115	4.068	< .001
Temperamental	0.428	0.123	3.490	< .001
Jealous	0.272	0.120	2.269	.023
Not contributing to the mean score				
Envious	−0.013	0.121	−0.109	.914
	Extraversion			
Mean score	0.164	0.087	1.875	.061
Contributing to the mean score				
Shy (recoded)	−0.126	0.113	−1.108	.268
Extroverted	0.207	0.111	1.866	.062
Talkative	0.223	0.096	2.326	.020
Lively	0.185	0.116	1.599	.110
Not contributing to the mean score				
Bashful (recoded)	0.071	0.129	0.555	.579
Quiet (recoded)	−0.076	0.116	−0.654	.513
Withdrawn (recoded)	−0.362	0.124	−2.927	.003
Enthusiastic	0.357	0.120	2.979	.003
	Conscientiousness			
Mean score	−0.036	0.100	−0.359	.720
Contributing to the mean score				
Inefficient (recoded)	−0.130	0.131	−0.988	.323
Careless (recoded)	−0.116	0.137	−0.845	.398
Sloppy (recoded)	−0.039	0.122	−0.316	.752
Disorganised (recoded)	−0.205	0.124	−1.652	.099
Orderly	0.358	0.112	3.182	.001
Not contributing to the mean score				
Efficient	0.288	0.121	2.393	.017
Systematic	0.501	0.124	4.039	< .001

Model	Estimate	*SE*	*t*	*p*
	Agreeableness			
Mean score	0.535	0.101	5.275	< .001
Contributing to the mean score				
Kind	0.675	0.124	5.454	< .001
Sympathetic	0.527	0.114	4.634	< .001
Cooperative	0.396	0.131	3.014	.003
Warm	0.504	0.124	4.078	< .001
Harsh (recoded)	−0.085	0.119	−0.717	.473
Not contributing to the mean score				
Cold (recoded)	−0.127	0.129	−0.979	.327
Selfish (recoded)	−0.277	0.123	−2.259	.024
	Openness			
Mean score	0.657	0.095	6.932	< .001
Contributing to the mean score				
Complex	0.423	0.115	3.671	< .001
Deep	0.545	0.115	4.725	< .001
Intellectual	0.260	0.107	2.421	.015
Philosophical	0.586	0.114	5.119	< .001
Imaginative	0.433	0.111	3.913	< .001
Not contributing to the mean score				
Creative	0.248	0.100	2.482	.013
Traditional (recoded)	−0.134	0.117	−1.147	.251

Note. The first wave was dummy-coded as 1 = first wave and 0 = all other waves. Parameters were estimated in separate fixed effects models with dummies for age. Panel-robust standard errors (as proposed by Arellano, 1987) were computed to account for serial correlation and heteroscedasticity.

Figure S1

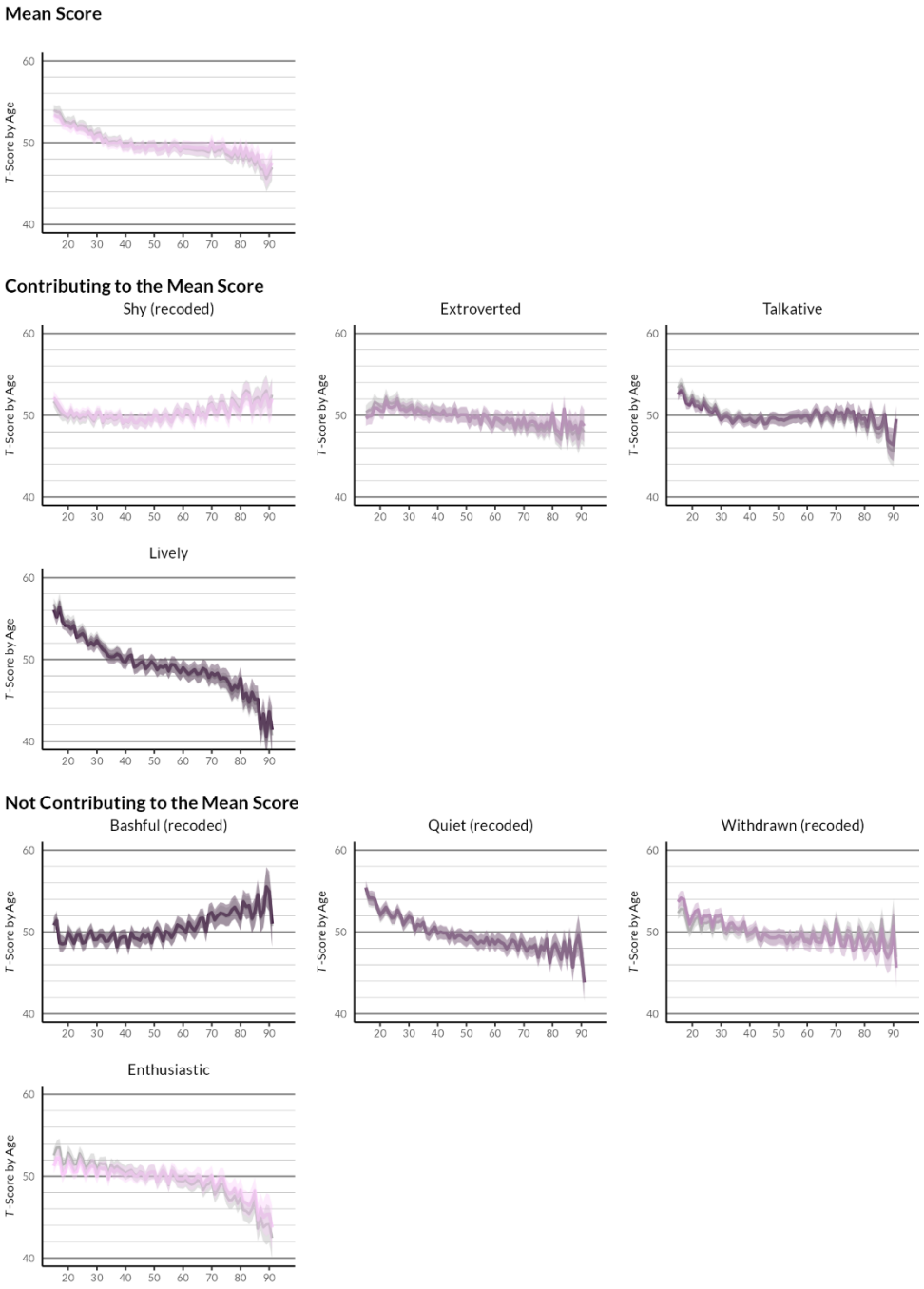
Within-Person Mean-Level Age Trajectories for Neuroticism in the Household, Income and Labour Dynamics in Australia (HILDA) Survey



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S2

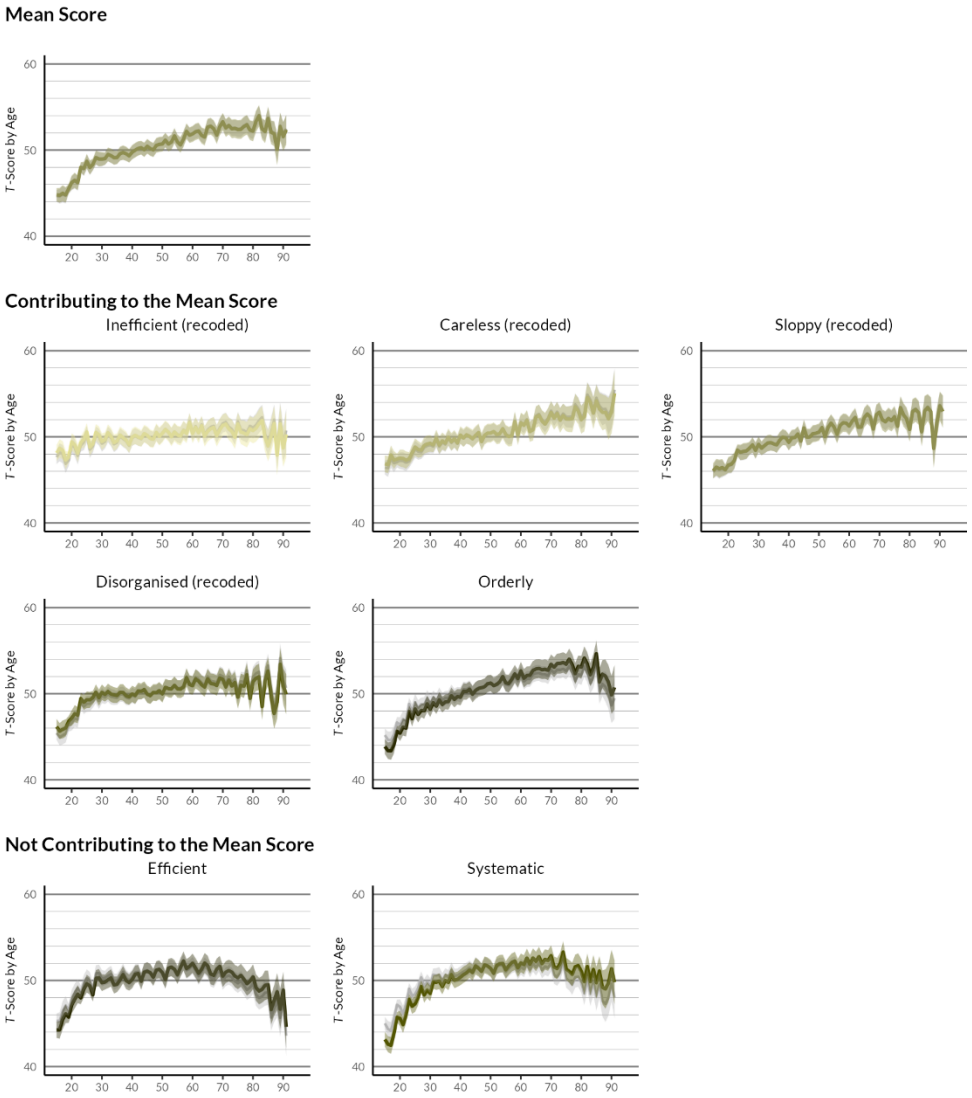
Within-Person Mean-Level Age Trajectories for Extraversion in the Household, Income, and Labour Dynamics in Australia (HILDA) Survey



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S3

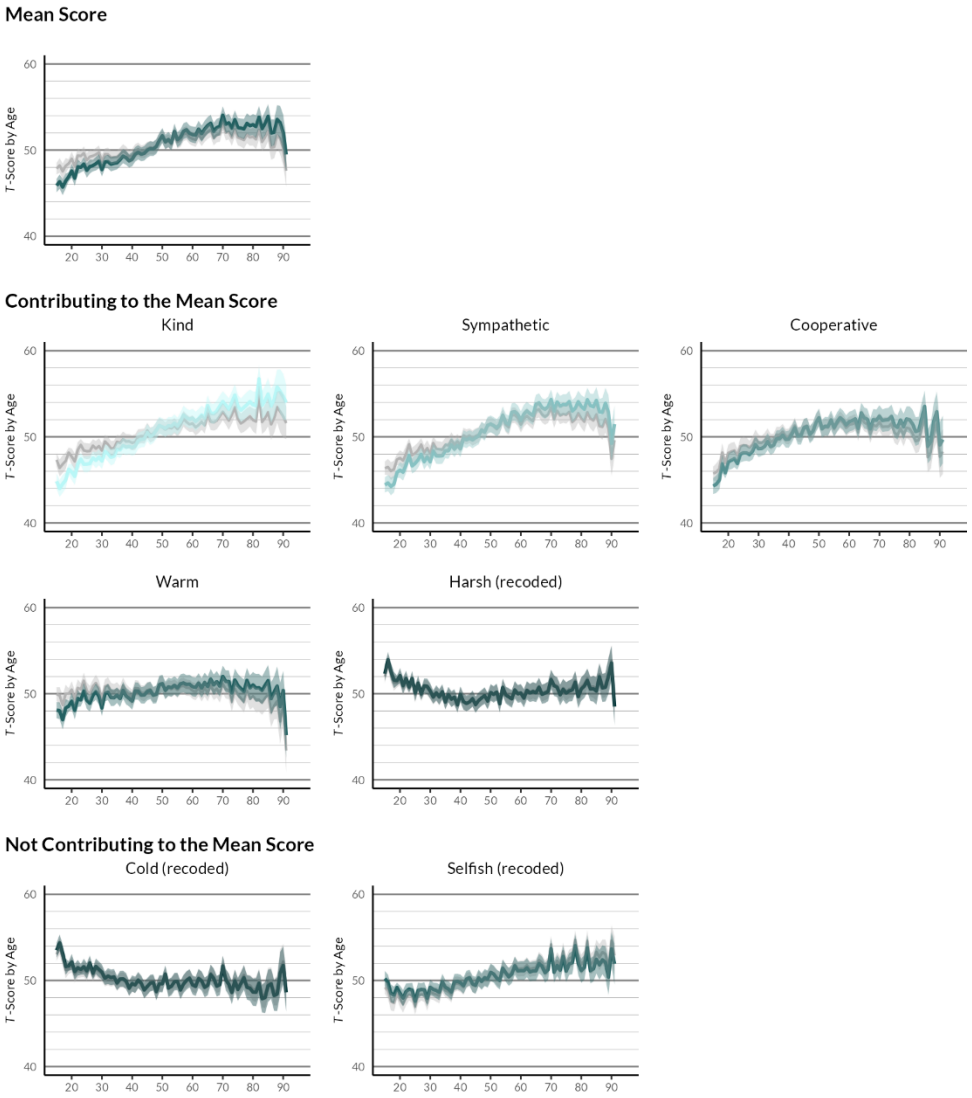
Within-Person Mean-Level Age Trajectories for Conscientiousness in the Household, Income and Labour Dynamics in Australia (HILDA) Survey



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S4

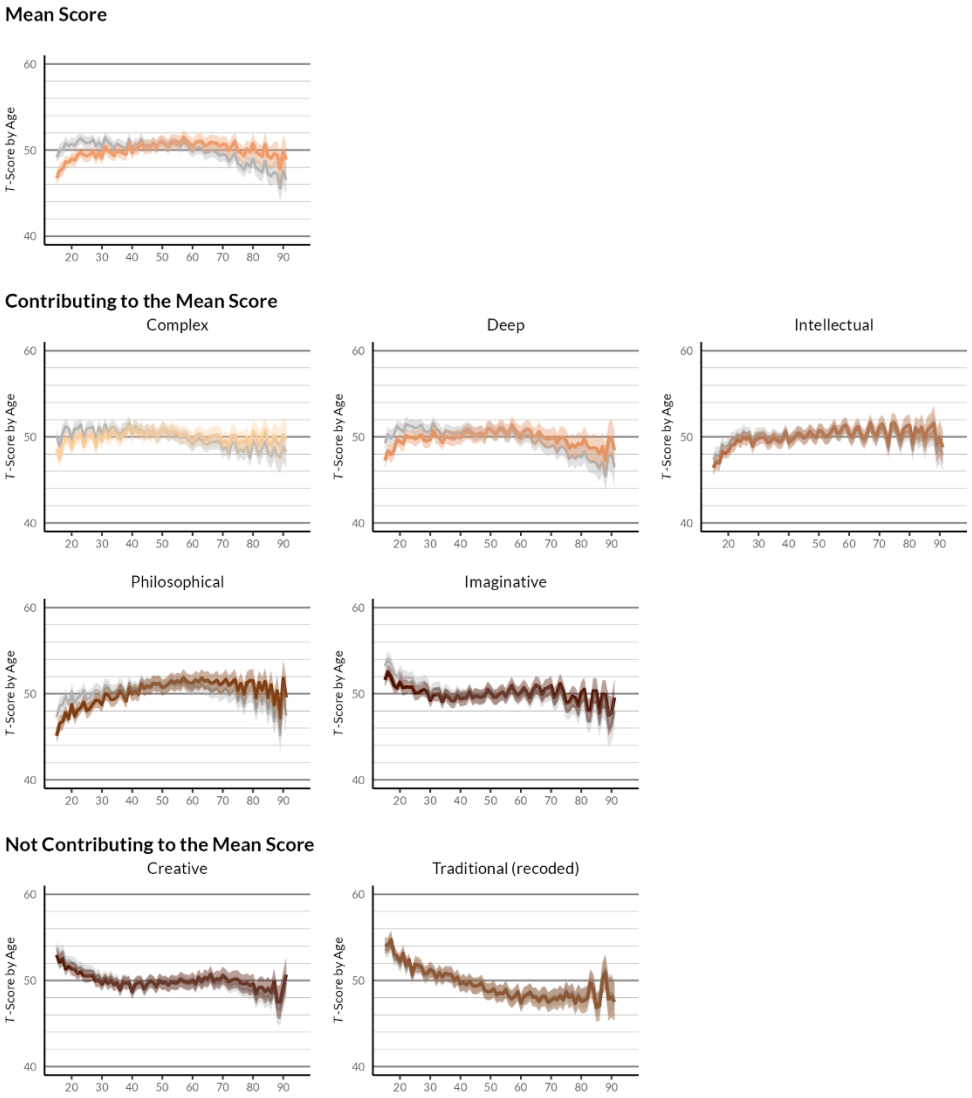
Within-Person Mean-Level Age Trajectories for Agreeableness in the Household, Income and Labour Dynamics in Australia (HILDA) Survey



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S5

Within-Person Mean-Level Age Trajectories for Openness in the Household, Income and Labour Dynamics in Australia (HILDA) Survey



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Supplemental Materials for Study 2: German Socio-Economic Panel (SOEP)

Sample Description and Psychometric Analyses

Table S5

Sample Characteristics by Trait in the Socio-Economic Panel (SOEP)

Trait	*n*	Age		% female
		M	*SD*	
Neuroticism	22,727	50.24	18.03	52.77
Extraversion	22,722	50.25	18.04	52.72
Conscientiousness	22,639	50.16	18.00	52.71
Agreeableness	22,716	50.25	18.04	52.73
Openness	22,593	50.20	18.01	52.69

Table S6

Overview of the Big Five Items in the Socio-Economic Panel (SOEP)

English (translated)	German (original)
Neuroticism	
Is relaxed, handles stress well [a]	Entspannt ist, mit Stress gut umgehen kann [a]
Gets nervous easily	Leicht nervös wird
Worries often	Sich oft Sorgen macht
Extraversion	
Is reserved [a]	Zurückhaltend ist [a]
Is communicative, talkative	Kommunikativ, gesprächig ist
Can be outgoing, is sociable	Aus sich herausgehen kann, gesellig ist
Conscientiousness	
Is rather lazy [a]	Eher faul ist [a]
Is effective and efficient in completing tasks	Aufgaben wirksam und effizient erledigt
Works thoroughly	Gründlich arbeitet
Agreeableness	
Is sometimes a bit rude to others [a]	Manchmal etwas grob zu anderen ist [a]
Is considerate and kind to others	Rücksichtsvoll und freundlich mit anderen umgeht
Can forgive	Verzeihen kann
Openness	
Values artistic, aesthetic experiences [b]	Künstlerische, ästhetische Erfahrungen schätzt [b]
Is original, comes up with new ideas	Originell ist, neue Ideen einbringt
Has a vivid fantasy, imagination	Eine lebhafte Phantasie, Vorstellungen hat
Is eager for knowledge [c]	Wissbegierig ist [c]

Note. Each item followed the statement "I am someone who …" ["Ich bin jemand, der …"].

[a] Recoded item. [b] In 2005, this item was "Values artistic experiences" ["Künstlerische Erfahrungen schätzt"]. [c] Not administered in 2005.

Table S7

Wave-Specific Psychometric Properties in the Socio-Economic Panel (SOEP)

Parameter	2005	2009	2013	2017	2019
			Neuroticism		
n	14,868	16,920	16,834	14,110	11,436
α	.60	.63	.62	.63	.66
ω	.63	.64	.65	.65	.68
			Extraversion		
n	14,870	16,893	16,831	14,107	11,419
α	.67	.67	.67	.69	.69
ω	.69	.69	.70	.71	.72
			Conscientiousness		
n	14,802	16,832	16,777	14,042	11,390
α	.66	.63	.63	.64	.64
ω	.68	.66	.66	.66	.67
			Agreeableness		
n	14,865	16,911	16,822	14,091	11,411
α	.54	.52	.52	.51	.53
ω	.59	.57	.56	.55	.58
			Openness [a]		
n	14,750	16,784	16,702	14,002	11,347
α	.63	.63	.62	.61	.62
ω	.64	.64	.63	.62	.63

[a] Based on items that were presented in each wave (i.e., “Is eager for knowledge” [“Wissbegierig ist”] was excluded, as it was not administered in 2005).

Fixed Effects Modeling

Table S8

Parameter Estimates for the Effects of the First Wave and Mode on Personality in the Socio-Economic Panel (SOEP)

Model	First wave				Mode			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
				Neuroticism				
Mean score	0.261	0.085	3.083	.002	1.140	0.124	9.166	< .001
Contributing to the mean score								
Is relaxed, handles stress well (recoded)	−0.435	0.097	−4.481	< .001	0.895	0.134	6.682	< .001
Gets nervous easily	0.152	0.095	1.600	.110	0.853	0.139	6.135	< .001
Worries often	0.803	0.094	8.522	< .001	0.843	0.136	6.191	< .001
				Extraversion				
Mean score	0.727	0.079	9.180	< .001	−0.911	0.116	−7.823	< .001
Contributing to the mean score								
Is reserved (recoded)	0.287	0.097	2.953	.003	0.055	0.139	0.392	.695
Is communicative, talkative	0.616	0.085	7.213	< .001	−1.046	0.123	−8.472	< .001
Can be outgoing, is sociable	0.815	0.090	9.042	< .001	−1.220	0.128	−9.523	< .001
				Conscientiousness				
Mean score	0.690	0.087	7.933	< .001	−1.167	0.132	−8.815	< .001
Contributing to the mean score								
Is rather lazy (recoded)	0.656	0.092	7.104	< .001	−0.924	0.138	−6.713	< .001
Is effective and efficient in completing tasks	0.456	0.101	4.498	< .001	−0.969	0.147	−6.602	< .001
Works thoroughly	0.388	0.095	4.064	< .001	−0.724	0.138	−5.258	< .001

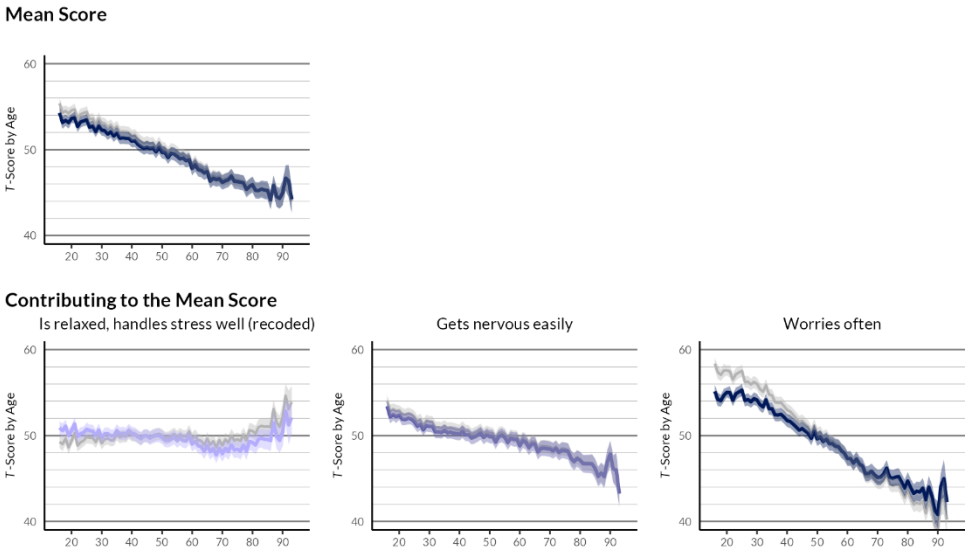
Model	First wave				Mode			
	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*t*	*p*
				Agreeableness				
Mean score	1.061	0.089	11.869	< .001	−1.172	0.133	−8.812	< .001
Contributing to the mean score								
Is sometimes a bit rude to others (recoded)	0.712	0.094	7.532	< .001	−1.075	0.140	−7.687	< .001
Is considerate and kind to others	0.884	0.101	8.761	< .001	−0.696	0.145	−4.794	< .001
Can forgive	0.710	0.102	6.935	< .001	−0.658	0.147	−4.488	< .001
				Openness				
Mean score	0.773	0.082	9.399	< .001	−0.755	0.117	−6.426	< .001
Contributing to the mean score								
Values artistic, aesthetic experiences	0.623	0.087	7.125	< .001	−0.739	0.125	−5.894	< .001
Is original, comes up with new ideas	0.412	0.094	4.366	< .001	−0.445	0.131	−3.398	< .001
Has a vivid fantasy, imagination	0.698	0.093	7.521	< .001	−0.485	0.130	−3.722	< .001
Not contributing to the mean score								
Is eager for knowledge [a]	−0.550	0.122	−4.522	< .001	−1.214	0.164	−7.417	< .001

Note. The first wave was dummy-coded as 1 = first wave and 0 = all other waves. Mode was dummy-coded as 1 = self-report questionnaire and 0 = interview. Self-administered questionnaires (SAQs) that were filled out without the help of the interviewer (either sent in by mail or collected by the interviewer) were classified as self-report questionnaires. Telephone interviews, oral interviews, SAQs filled out with the help of the interviewer, oral interviews combined with SAQs, and computer-assisted personal interviews (CAPIs) were classified as interviews. Parameters for the first wave and mode were estimated simultaneously in a fixed effects model with dummies for age, separately for the mean score and each item. Panel-robust standard errors (as proposed by Arellano, 1987) were computed to account for serial correlation and heteroscedasticity.

[a] n = 17,718 with 91 as the oldest age (item was not administered in 2005).

Figure S6

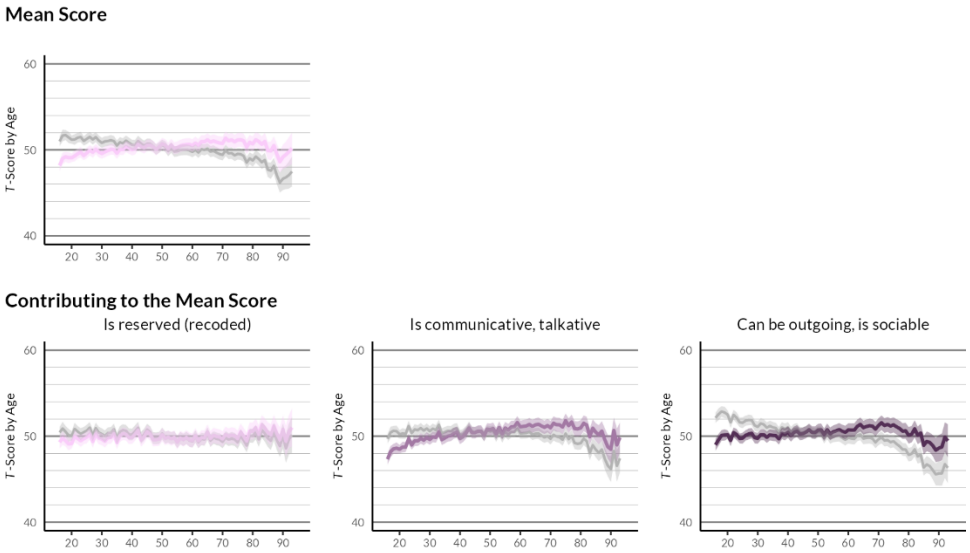
Within-Person Mean-Level Age Trajectories for Neuroticism in the Socio-Economic Panel (SOEP)



Note. The grey line indicates the trajectory when the first wave and mode were not controlled for. The shaded area represents the panel-robust standard error.

Figure S7

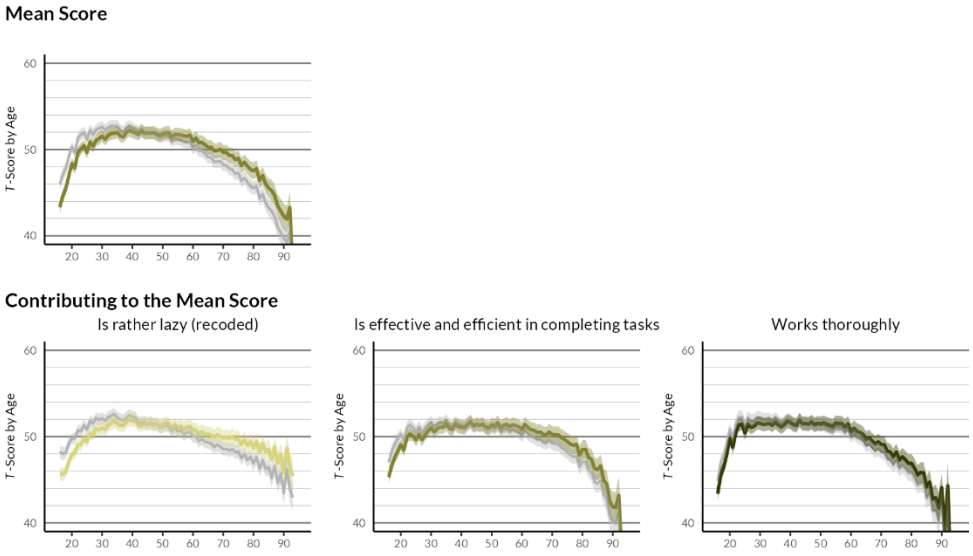
Within-Person Mean-Level Age Trajectories for Extraversion in the Socio-Economic Panel (SOEP)



Note. The grey line indicates the trajectory when the first wave and mode were not controlled for. The shaded area represents the panel-robust standard error.

Figure S8

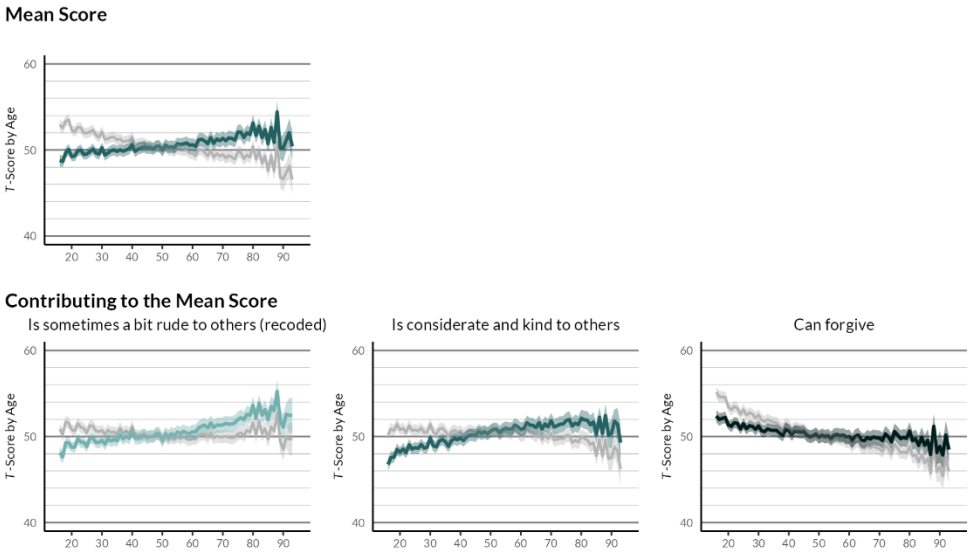
Within-Person Mean-Level Age Trajectories for Conscientiousness in the Socio-Economic Panel (SOEP)



Note. The grey line indicates the trajectory when the first wave and mode were not controlled for. The shaded area represents the panel-robust standard error.

Figure S9

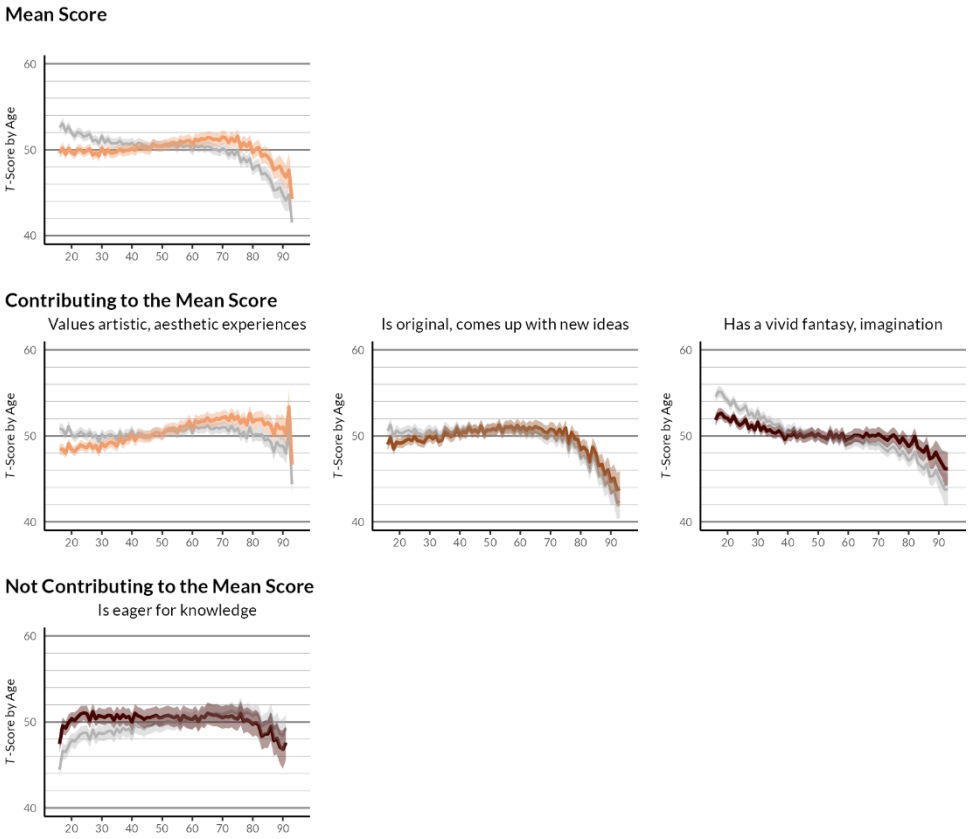
Within-Person Mean-Level Age Trajectories for Agreeableness in the Socio-Economic Panel (SOEP)



Note. The grey line indicates the trajectory when the first wave and mode were not controlled for. The shaded area represents the panel-robust standard error.

Figure S10

Within-Person Mean-Level Age Trajectories for Openness in the Socio-Economic Panel (SOEP)



Note. The grey line indicates the trajectory when the first wave and mode were not controlled for. The shaded area represents the panel-robust standard error.

Supplemental Materials for Study 3: Dutch Longitudinal Internet Studies for the Social Sciences (LISS) Panel

Item Selection

All items in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel that were used to assess the Big Five are presented in Table S9. Participants were included in the item selection analyses if they answered at least two items from the respective Big Five personality trait within one wave. The item selection procedure was performed by means of confirmatory factor analysis (CFA) separately for each wave and personality factor. Prior to the analyses, reversed items were recoded so that higher values indicated higher construct expression. In a CFA, all items constituted the manifest indicators of one latent factor (i.e., the corresponding personality construct in one wave). No residual correlations were modeled; no additional constraints were added to the measurement model. Comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) were used as fit indices. Full information maximum likelihood estimation (FIML) was used to account for missing data.

If a CFA model consistently did not achieve an acceptable model fit, in the next step, one item was excluded. The cutoffs by Hu and Bentler (1999) were used to evaluate the level of model fit (i.e., for an acceptable model fit, CFI $\geq$ .95, RMSEA $\leq$ .06, and SRMR $\leq$ .08). We iteratively repeated this procedure until an acceptable fit was achieved. Choosing an item for exclusion was based on semantic as well as statistical criteria, especially factor loadings and modification indices (MIs). To avoid overfitting, we only considered statistical features that emerged consistently across several waves. This approach is similar to Seifert et al.'s (2022) item

selection procedure for the Big Five measure in the Household, Income and Labour Dynamics in Australia (HILDA) Survey.

The lavaan package (Rosseel, 2012) was used to estimate the CFA models. We computed α and ω using the semTools package (Jorgensen et al., 2021). All analysis scripts (including more detailed results) are publicly available on the Open Science Framework and can be accessed at https://osf.io/8rjex/.

Table S9

Overview of the Big Five Items in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel

English (translated)	Dutch (original)
Neuroticism	
Get stressed out easily	Raak makkelijk gestresst
Am relaxed most of the time [a]	Ben meestal ontspannen [a]
Worry about things	Maak me zorgen over dingen
Seldom feel blue [a]	Voel me zelden neerslachtig [a]
Am easily disturbed [b]	Ben snel verontrust [b]
Get upset easily	Raak makkelijk van streek
Change my mood a lot [b]	Verander vaak van stemming [b]
Have regular mood swings [b, c]	Heb regelmatig stemmingswisselingen [b]
Get irritated easily	Raak snel geërgerd
Often feel blue [b]	Voel me vaak neerslachtig [b]
Extraversion	
Liven things up [d]	Breng leven in de brouwerij
Don't talk a lot [a]	Praat niet veel [a]
Feel comfortable around people	Voel me goed in het gezelschap van mensen
Keep in the background [a, b]	Blijf op de achtergrond [a, b]
Start conversations	Begin gesprekken
Have little to say [a, b]	Heb weinig te zeggen [a, b]
Talk to a lot of different people at parties [b]	Praat met veel verschillende mensen op feestjes [b]
Don't like to draw attention to myself [a, b]	Houd er niet van de aandacht op mijzelf te vestigen [a, b]
Don't mind being the center of attention	Vind het niet erg om in het middelpunt van de belangstelling te staan
Am quiet around strangers [a]	Ben stil in het gezelschap van vreemden [a]
Conscientiousness	
Am always prepared [b]	Ben altijd voorbereid [b]
Leave my belongings around [a, b]	Laat mijn persoonlijke bezittingen rondslingeren [a, b]
Pay attention to details	Besteed aandacht aan details
Make a mess of things [a, b]	Maak een puinhoop van dingen [a, b]
Get chores done right away	Doe karweitjes meteen
Often forget to put things back in their proper place [a]	Vergeet vaak om dingen op de juiste plaats terug te zetten [a]
Like order	Houd van orde
Shirk my duties [a]	Onttrek me aan mijn verplichtingen [a]
Follow a schedule	Volg een planning
Am exacting in my work [b]	Ben veeleisend in mijn werk [b]

English (translated)	Dutch (original)
Agreeableness	
Feel little concern for others [a]	Voel me weinig begaan met anderen [a]
Am interested in people [b]	Ben geïnteresseerd in mensen [b]
Insult people [a]	Beledig mensen [a]
Sympathize with others' feelings	Voel mee met de gevoelens van anderen
Am not interested in other people's problems [a, b]	Ben niet geïnteresseerd in de problemen van andere mensen [a, b]
Am sentimental [b, e]	Ben sentimenteel [b]
Am not really interested in others [a]	Ben niet echt geïnteresseerd in anderen [a]
Take time out for others	Neem de tijd voor anderen
Feel others' emotions [b]	Voel de emoties van anderen [b]
Make people feel at ease	Zorg dat mensen zich op hun gemak voelen
Openness	
Have a rich vocabulary	Heb een uitgebreide woordenschat
Have difficulty understanding abstract ideas [a, b]	Heb moeite om abstracte ideeën te begrijpen [a, b]
Have a vivid imagination [b]	Heb een levendige fantasie [b]
Am not interested in abstract ideas [a]	Ben niet geïnteresseerd in abstracte ideeën [a]
Have excellent ideas [b]	Heb uitstekende ideeën [b]
Do not have a good imagination [a]	Heb geen goede verbeelding [a]
Am quick to understand things	Begrijp dingen snel
Use difficult words [b]	Gebruik moeilijke woorden [b]
Spend time reflecting on things	Besteed tijd om over dingen na te denken
Am full of ideas	Zit vol met ideeën

Note. Items stem from the 50-item International Personality Item Pool (IPIP) Version of the Goldberg (1992) markers for the Big Five personality traits.

[a] Recoded item. [b] Excluded item. [c] Original IPIP wording was "Have frequent mood swings." [d] Original IPIP wording was "Am the life of the party." [e] Original IPIP wording was "Have a soft heart."

Neuroticism

Measuring Neuroticism with 10 items yielded unsatisfactory psychometric criteria (see Table S10). For each wave, the highest MI indicated a residual correlation between the items "Change my mood a lot" and "Have regular mood swings," possibly due to overlapping content. The second highest MI in each wave was considerably smaller. Excluding either of the two items led to a considerable improvement in model fit (with comparable fit indices and internal consistencies). However, in each wave, the item "Have regular mood swings" had a higher factor loading than "Change my mood a lot." Hence, we excluded "Change my mood a lot."

Afterwards, further improvement in the psychometric criteria was still indicated (see Table S10). With nine items, the highest MI for each wave indicated a residual correlation between the items "Have regular mood swings" and "Often feel blue" (the second highest MI was considerably smaller), a result that was potentially due to content similarity. Excluding "Often feel blue" led to more pronounced improvements in model fit than excluding "Have regular mood swings." In addition, "Often feel blue" is the opposite of the Neuroticism item "Seldom feel blue." Accordingly, "Often feel blue" was excluded.

Model fit improved but not on a satisfactory level (see Table S10). Amongst the highest MIs, a residual correlation between the items "Worry about things" and "Am easily disturbed" was consistently indicated, mirroring the semantic intersection of the two items (this residual correlation yielded the highest MI especially in the waves in which the whole sample provided personality data). Excluding "Am easily disturbed" went along with a greater improvement in model fit than excluding "Worry about things." Hence, "Am easily disturbed" was excluded.

The model fit with seven items could still be improved (see Table S10). Among the highest MIs, a residual correlation between “Have regular mood swings” and “Get irritated easily” was indicated, a result that might be explained by a related content. Without “Have regular mood swings,” the model fit was better than without “Get irritated easily.” Accordingly, “Have regular mood swings” was excluded, which yielded a good model fit (see Table S10).

Table S10

Psychometric Criteria for the Item Selection for Neuroticism

Wave	*n*	$\chi^{2\,a}$	CFI	RMSEA	SRMR	α	ω
All items (*df* = 35)							
2008	6,773	4,081.31	.854	.131	.058	.88	.88
2009	5,619	3,644.13	.848	.135	.058	.88	.88
2010	1,375	825.35	.843	.128	.059	.86	.86
2011	5,292	3,198.03	.862	.131	.055	.88	.89
2012	1,476	888.83	.852	.129	.058	.87	.88
2013	5,139	2,825.94	.872	.125	.053	.88	.88
2014	6,512	4,009.46	.863	.132	.055	.89	.89
2015	506	366.17	.866	.137	.057	.90	.90
2017	6,062	3,793.26	.860	.133	.056	.89	.89
2018	807	507.49	.867	.129	.057	.89	.89
2019	5,037	3,057.76	.875	.131	.052	.90	.90
2020	5,890	3,942.61	.863	.138	.054	.89	.90
Excluding “Change my mood a lot” (*df* = 27)							
2008	6,773	2,491.88	.890	.116	.049	.86	.86
2009	5,619	2,188.46	.887	.119	.049	.86	.87
2010	1,375	462.47	.891	.108	.048	.84	.85
2011	5,292	2,112.65	.889	.121	.049	.87	.87
2012	1,476	575.42	.885	.117	.052	.86	.86
2013	5,139	1,730.68	.906	.111	.045	.87	.87
2014	6,512	2,380.28	.902	.116	.045	.88	.88
2015	506	229.68	.901	.122	.050	.89	.89
2017	6,062	2,243.95	.900	.116	.047	.87	.88
2018	807	275.92	.916	.107	.045	.88	.88
2019	5,037	1,843.87	.909	.116	.045	.88	.89
2020	5,890	2,085.60	.911	.114	.044	.88	.89
Excluding “Often feel blue” (*df* = 20)							
2008	6,773	1,051.29	.939	.087	.037	.84	.84
2009	5,619	929.21	.937	.090	.038	.84	.84
2010	1,375	206.17	.936	.082	.038	.81	.82
2011	5,292	939.93	.935	.093	.039	.85	.85
2012	1,476	311.37	.923	.099	.042	.84	.84
2013	5,139	690.45	.952	.081	.034	.85	.85
2014	6,512	1,096.40	.942	.091	.036	.85	.86
2015	506	109.13	.945	.094	.040	.87	.87
2017	6,062	1,045.31	.940	.092	.038	.85	.85
2018	807	128.37	.953	.082	.036	.85	.86
2019	5,037	888.47	.944	.093	.037	.86	.87
2020	5,890	931.93	.950	.088	.035	.86	.87

Wave	*n*	χ^2 [a]	CFI	RMSEA	SRMR	α	ω
Excluding "Am easily disturbed" (*df* = 14)							
2008	6,773	588.73	.956	.078	.030	.81	.82
2009	5,619	528.51	.953	.081	.032	.81	.82
2010	1,375	128.99	.949	.077	.034	.78	.79
2011	5,292	553.94	.950	.085	.033	.82	.82
2012	1,476	208.84	.932	.097	.039	.81	.82
2013	5,139	439.38	.960	.077	.031	.82	.82
2014	6,512	710.06	.951	.087	.032	.83	.83
2015	506	64.32	.962	.084	.032	.86	.86
2017	6,062	624.85	.954	.085	.032	.83	.83
2018	807	106.80	.948	.091	.038	.83	.83
2019	5,037	567.28	.954	.089	.032	.84	.84
2020	5,890	602.69	.958	.084	.031	.84	.84
Excluding "Have regular mood swings" (*df* = 9)							
2008	6,773	196.91	.981	.056	.021	.78	.79
2009	5,619	193.40	.979	.060	.024	.78	.79
2010	1,375	48.97	.977	.057	.026	.75	.76
2011	5,292	278.72	.967	.075	.029	.79	.79
2012	1,476	92.37	.963	.079	.031	.79	.79
2013	5,139	240.29	.972	.071	.027	.79	.80
2014	6,512	268.92	.976	.067	.025	.80	.81
2015	506	30.31	.979	.068	.026	.83	.84
2017	6,062	277.99	.974	.070	.026	.80	.80
2018	807	52.17	.969	.077	.031	.80	.81
2019	5,037	275.27	.972	.077	.028	.81	.82
2020	5,890	240.47	.979	.066	.024	.81	.82

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation;

SRMR = standardized root mean square residual.

[a] All *p*s < .001.

Extraversion

Measuring Extraversion with 10 items did not provide acceptable fit indices (see Table S11). In a first step, we excluded "Talk to a lot of different people at parties" as the contextualization of the item (i.e., being talkative *at parties*) is not appropriate for assessing talkativeness across the entire life span (especially in older ages; see Olaru et al., 2019).

Excluding this item improved the model fit but not to a satisfactory level (see Table S11). In each wave, a residual correlation between the items "Don't like to draw attention to myself" and "Don't mind being the center of attention" was indicated by the MIs. The second highest MI was considerably smaller. The two items are opposites, which might explain this pattern of results. Except for a single wave, "Don't like to draw attention to myself" had the lowest factor loading of all the items. In addition, excluding "Don't like to draw attention to myself" always yielded better fit criteria and in most cases higher internal consistencies than excluding "Don't mind being the center of attention." Lastly, "Don't like to draw attention to myself" might be misinterpreted by respondents as capturing low self-awareness instead of Extraversion. For these reasons, we excluded "Don't like to draw attention to myself."

Without this item, the model fit improved, but there was room for further improvement (see Table S11). A residual correlation between the items "Keep in the background" and "Start conversations" as well as between the items "Keep in the background" and "Don't mind being the center of attention" was each among the highest MIs. We decided to investigate the residual correlation between "Keep in the background" and "Don't mind being the center of attention" in more detail, as it seemed more plausible that this item pair would be correlated; the items are opposites, and they share a figurative language. As the model fit usually improved more without

"Keep in the background" than without "Don't mind being the center of attention," we excluded "Keep in the background."

The model fit improved when this item was excluded (see Table S11). Now, a residual correlation between the items "Don't talk a lot" and "Have little to say" was indicated by the highest MIs. The two items cover a similar aspect (and both are recoded). Excluding "Have little to say" led to higher internal consistencies than excluding "Don't talk a lot," whereas the fit indices did not differ substantially between the two options. Hence, we excluded "Have little to say," resulting in a final model with good fit (see Table S11).

Table S11

Psychometric Criteria for the Item Selection for Extraversion

Wave	*n*	χ^{2} [a]	CFI	RMSEA	SRMR	α	ω
All items (*df* = 35)							
2008	6,773	2,329.46	.900	.098	.049	.86	.86
2009	5,619	2,330.62	.888	.108	.052	.87	.87
2010	1,375	537.31	.880	.102	.053	.85	.85
2011	5,292	2,264.21	.882	.110	.054	.87	.87
2012	1,476	566.14	.903	.101	.049	.87	.88
2013	5,139	1,936.39	.902	.103	.050	.87	.88
2014	6,512	2,911.51	.886	.112	.054	.88	.88
2015	506	287.94	.860	.120	.062	.86	.87
2017	6,062	2,500.91	.894	.108	.052	.88	.88
2018	807	394.92	.887	.113	.053	.88	.88
2019	5,037	2,339.03	.886	.114	.054	.88	.88
2020	5,890	2,612.44	.894	.112	.051	.88	.89
Excluding “Talk to a lot of different people at parties” (*df* = 27)							
2008	6,773	1,616.70	.914	.093	.045	.84	.85
2009	5,619	1,631.49	.904	.103	.049	.85	.85
2010	1,375	350.25	.902	.093	.048	.82	.83
2011	5,292	1,619.45	.897	.106	.050	.85	.85
2012	1,476	395.45	.918	.096	.046	.86	.86
2013	5,139	1,417.17	.912	.100	.048	.86	.86
2014	6,512	2,119.87	.899	.109	.051	.86	.86
2015	506	213.66	.875	.117	.060	.85	.85
2017	6,062	1,791.18	.908	.104	.049	.86	.86
2018	807	296.37	.899	.111	.050	.86	.87
2019	5,037	1,651.45	.903	.109	.051	.86	.87
2020	5,890	1,911.35	.906	.109	.050	.87	.87
Excluding “Don’t like to draw attention to myself” (*df* = 20)							
2008	6,773	627.40	.962	.067	.028	.83	.84
2009	5,619	634.83	.957	.074	.030	.85	.85
2010	1,375	154.00	.953	.070	.032	.82	.82
2011	5,292	691.66	.950	.080	.033	.84	.85
2012	1,476	145.60	.967	.065	.027	.85	.85
2013	5,139	623.30	.957	.077	.031	.85	.85
2014	6,512	840.79	.954	.079	.031	.85	.86
2015	506	77.53	.954	.075	.035	.84	.84
2017	6,062	761.48	.955	.078	.032	.85	.86
2018	807	152.48	.940	.091	.038	.85	.86
2019	5,037	718.91	.951	.083	.033	.86	.86
2020	5,890	807.81	.955	.082	.032	.86	.87

Wave	*n*	χ^2 [a]	CFI	RMSEA	SRMR	α	ω
Excluding "Keep in the background" (*df* = 14)							
2008	6,773	301.28	.977	.055	.023	.81	.81
2009	5,619	308.45	.973	.061	.024	.82	.82
2010	1,375	92.27	.963	.064	.029	.78	.79
2011	5,292	353.76	.967	.068	.028	.82	.82
2012	1,476	63.51	.983	.049	.021	.82	.83
2013	5,139	331.54	.970	.066	.027	.83	.83
2014	6,512	385.55	.973	.064	.025	.83	.83
2015	506	47.89	.964	.069	.032	.81	.81
2017	6,062	462.67	.965	.073	.029	.83	.83
2018	807	82.18	.960	.078	.032	.83	.83
2019	5,037	377.06	.967	.072	.028	.83	.83
2020	5,890	424.98	.969	.071	.028	.84	.84
Excluding "Have little to say" (*df* = 9)							
2008	6,773	167.43	.984	.051	.019	.78	.79
2009	5,619	179.97	.980	.058	.021	.79	.80
2010	1,375	55.84	.972	.062	.025	.75	.76
2011	5,292	194.53	.977	.062	.022	.79	.80
2012	1,476	44.10	.985	.051	.020	.80	.81
2013	5,139	210.02	.977	.066	.023	.80	.81
2014	6,512	216.27	.981	.059	.020	.81	.81
2015	506	25.27 [b]	.978	.060	.027	.79	.79
2017	6,062	282.69	.973	.071	.025	.81	.81
2018	807	42.11	.976	.068	.025	.81	.81
2019	5,037	194.27	.979	.064	.022	.81	.81
2020	5,890	268.55	.975	.070	.024	.82	.82

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation;

SRMR = standardized root mean square residual.

[a] All *p*s < .001 (if not indicated otherwise). [b] $p = .003$.

Conscientiousness

As Table S12 shows, assessing Conscientiousness without excluding any items returned an unacceptable fit. In each wave, the item "Am exacting in my work" showed a markedly low factor loading with an average loading of .278 across waves. In a first step, we excluded this item.

Excluding "Am exacting in my work" improved the fit but not to a satisfactory level (see Table S12). After that, the highest MI indicated a residual correlation between the item "Leave my belongings around" and "Often forget to put things back in their proper place" or a residual correlation between the item "Make a mess of things" and "Shirk my duties." We took a closer look at the residual correlation between "Leave my belongings around" and "Often forget to put things back in their proper place," as this residual correlation was more clearly due to a semantic overlap. Predominantly, the factor loading of "Often forget to put things back in their proper place" was higher than the factor loading of "Leave my belongings around." Excluding "Leave my belongings around" consistently resulted in a better model fit and mostly higher internal consistencies than excluding "Often forget to put things back in their proper place." Accordingly, "Leave my belongings around" was excluded.

Despite the fact that the model fit improved after the item was removed, an acceptable level was not achieved (see Table S12). Now, for each wave, the highest MI indicated a residual correlation between "Make a mess of things" and "Shirk my duties." Typically, the fit improved to a greater extent when "Make a mess of things" was excluded instead of "Shirk my duties." In addition, the content of "Make a mess of things" is also reflected by "Often forget to put things back in their proper place," indicating a redundancy in the items. Hence, we excluded "Make a mess of things."

After the exclusion of that item, the model showed a better fit, but it seemed possible to improve the model further (see Table S12). The items “Am always prepared” and “Follow a schedule” exhibit a semantic overlap; and, not surprisingly, the highest MI indicated a residual correlation between the two items in the majority of waves. With one exception, the factor loading of “Follow a schedule” was higher than of “Am always prepared.” Further, excluding “Am always prepared” went along with higher internal consistencies than excluding “Follow a schedule.” Similarly, higher estimates of the CFI and lower estimates of the SRMR were observed when “Am always prepared” was excluded instead of “Follow a schedule” (differences in the RMSEA were weak and inconclusive). In line with these points, “Am always prepared” was excluded, resulting in an overall model fit that was acceptable to good (see Table S12).

Table S12

Psychometric Criteria for the Item Selection for Conscientiousness

Wave	*n*	$\chi^{2\,a}$	CFI	RMSEA	SRMR	α	ω
All items (*df* = 35)							
2008	6,773	2,129.64	.838	.094	.056	.77	.78
2009	5,619	1,828.34	.850	.095	.055	.79	.80
2010	1,375	568.44	.817	.105	.065	.78	.79
2011	5,292	1,970.62	.828	.102	.060	.79	.79
2012	1,476	581.72	.822	.103	.061	.78	.79
2013	5,139	1,861.73	.828	.101	.060	.78	.79
2014	6,512	2,283.50	.833	.099	.058	.78	.79
2015	506	272.44	.816	.116	.066	.81	.82
2017	6,062	2,314.31	.812	.104	.061	.78	.78
2018	807	262.67	.856	.090	.055	.78	.79
2019	5,037	1,853.91	.825	.102	.060	.78	.79
2020	5,889	2,297.05	.822	.105	.061	.79	.79
Excluding “Am exacting in my work” (*df* = 27)							
2008	6,773	1,490.58	.877	.089	.048	.77	.78
2009	5,619	1,329.58	.882	.093	.048	.79	.80
2010	1,375	406.54	.859	.101	.056	.79	.79
2011	5,292	1,502.34	.859	.102	.055	.79	.80
2012	1,476	441.34	.854	.102	.055	.78	.79
2013	5,139	1,357.47	.864	.098	.052	.79	.79
2014	6,512	1,669.73	.868	.097	.051	.79	.79
2015	506	211.04	.846	.116	.060	.81	.82
2017	6,062	1,767.43	.845	.103	.056	.78	.79
2018	807	215.08	.875	.093	.053	.79	.79
2019	5,037	1,446.75	.854	.102	.055	.79	.79
2020	5,889	1,762.63	.854	.104	.055	.79	.80
Excluding “Leave my belongings around” (*df* = 20)							
2008	6,769	910.08	.895	.081	.041	.73	.74
2009	5,619	856.17	.897	.086	.043	.76	.76
2010	1,374	254.29	.878	.092	.050	.75	.75
2011	5,287	1,005.18	.871	.097	.049	.75	.76
2012	1,475	234.34	.896	.085	.045	.75	.76
2013	5,135	985.22	.867	.097	.049	.75	.76
2014	6,507	1,129.08	.880	.092	.046	.75	.76
2015	505	128.47	.878	.104	.052	.79	.79
2017	6,061	1,206.48	.858	.099	.051	.75	.75
2018	806	148.60	.886	.089	.049	.75	.76
2019	5,034	987.12	.866	.098	.050	.76	.76
2020	5,884	1,146.66	.871	.098	.049	.76	.76

Wave	*n*	χ^2 [a]	CFI	RMSEA	SRMR	α	ω
Excluding "Make a mess of things" (*df* = 14)							
2008	6,769	402.03	.937	.064	.031	.70	.70
2009	5,619	347.44	.945	.065	.031	.72	.73
2010	1,374	133.46	.918	.079	.040	.71	.72
2011	5,287	480.68	.918	.079	.037	.72	.73
2012	1,475	115.75	.931	.070	.035	.71	.72
2013	5,135	394.24	.929	.073	.035	.72	.72
2014	6,507	443.05	.937	.069	.032	.72	.73
2015	505	50.53	.942	.072	.036	.75	.76
2017	6,061	477.09	.925	.074	.035	.72	.72
2018	806	81.62	.923	.077	.039	.73	.73
2019	5,034	359.85	.935	.070	.033	.72	.73
2020	5,884	469.12	.929	.074	.035	.72	.73
Excluding "Am always prepared" (*df* = 9)							
2008	6,768	237.57	.956	.061	.026	.68	.69
2009	5,617	235.71	.956	.067	.028	.71	.72
2010	1,372	88.13	.935	.080	.036	.69	.70
2011	5,285	266.73	.945	.074	.030	.70	.71
2012	1,474	68.59	.950	.067	.030	.68	.69
2013	5,135	227.84	.950	.069	.028	.70	.71
2014	6,504	283.02	.951	.068	.029	.70	.71
2015	503	24.90 [b]	.968	.059	.028	.73	.73
2017	6,057	294.25	.943	.072	.030	.69	.70
2018	804	48.81	.938	.074	.034	.69	.69
2019	5,033	226.95	.949	.069	.029	.70	.70
2020	5,884	239.11	.954	.066	.028	.70	.70

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation;

SRMR = standardized root mean square residual.

[a] All *p*s < .001 (if not indicated otherwise). [b] $p = .003$.

Agreeableness

As shown in Table S13, assessing Agreeableness with 10 items consistently returned improvable psychometric criteria. Strikingly, “Am sentimental” showed low factor loadings in each wave with an average loading of .228. As an explanation for this result, emotionality might be covered better by Neuroticism than Agreeableness (see John et al., 2008). Correspondingly, “Am sentimental” was excluded.

The model fit improved when only nine items were included but not to a sufficient degree (see Table S13). Now, the highest MI indicated a residual correlation between “Sympathize with others’ feelings” and “Feel others’ emotions,” and with the same frequency, there was also a residual correlation between “Am not interested in other people’s problems” and “Am not really interested in others.” First, the residual correlation between “Sympathize with others’ feelings” and “Feel others’ emotions” was investigated, as this residual correlation would still be indicated by the highest MI if either “Am not interested in other people’s problems” or “Am not really interested in others” were excluded first. The residual correlation between “Sympathize with others’ feelings” and “Feel others’ emotions” might be explained by their overlapping content. In each wave, “Sympathize with others’ feelings” had a higher factor loading than “Feel others’ emotions.” In addition, excluding “Feel others’ emotions” led to higher internal consistencies (in each wave) and better improvements in model fit (with exception of a single wave) than the exclusion of “Sympathize with others’ feelings.” Further, “Feel others’ emotions” might be somewhat ambiguously understood in the sense of whether the statement is directed toward one’s *compassion* with or one’s *awareness* of others’ emotions. Accordingly, “Feel others’ emotions” was excluded.

Without this item, the model fit improved (see Table S13), and the highest MIs quasi-equivocally indicated a residual correlation between “Am not interested in other people’s problems” and “Am not really interested in others” (potentially explained by their semantic overlap). In each wave, the item “Am not really interested in others” showed a higher factor loading than “Am not interested in other people’s problems.” In addition, the internal consistencies were consistently higher when “Am not interested in other people’s problems” instead of “Am not really interested in others” was excluded (fit indices were inconclusive). Lastly, “Am not interested in other people’s problems” reflects specific point of “Am not really interested in others.” Following these points, “Am not interested in other people’s problems” was excluded.

Without this item, the model fit improved (see Table S13). Unfavorably, among the remaining items “Am interested in people” is the very opposite of “Am not really interested in others.” To avoid redundancy and to keep a balance between recoded and nonrecoded items, we decided to exclude “Am interested in people.” Thereafter, we found an overall good model fit (see Table S13).

Table S13

Psychometric Criteria for the Item Selection for Agreeableness

Wave	*n*	χ^2 [a]	CFI	RMSEA	SRMR	α	ω
All items (*df* = 35)							
2008	6,773	1,335.83	.921	.074	.039	.80	.80
2009	5,619	1,080.99	.932	.073	.037	.81	.82
2010	1,375	365.46	.891	.083	.050	.77	.77
2011	5,292	916.52	.937	.069	.035	.81	.82
2012	1,476	307.80	.920	.073	.040	.78	.79
2013	5,139	1,013.98	.929	.074	.038	.81	.82
2014	6,511	1,333.00	.928	.075	.038	.81	.82
2015	506	118.71	.942	.069	.042	.83	.83
2017	6,062	1,547.40	.912	.084	.043	.82	.82
2018	807	255.53	.902	.088	.046	.82	.82
2019	5,037	1,103.51	.928	.078	.039	.82	.83
2020	5,890	1,480.01	.918	.084	.041	.83	.83
Excluding “Am sentimental” (*df* = 27)							
2008	6,773	1,071.31	.934	.076	.037	.82	.82
2009	5,619	834.73	.946	.073	.033	.83	.84
2010	1,375	290.15	.911	.084	.046	.80	.80
2011	5,292	688.18	.951	.068	.031	.83	.83
2012	1,476	248.72	.933	.075	.037	.81	.81
2013	5,139	767.34	.945	.073	.033	.83	.84
2014	6,511	1,081.31	.939	.077	.035	.84	.84
2015	506	83.17	.959	.064	.035	.84	.85
2017	6,062	1,258.21	.926	.087	.039	.84	.84
2018	807	219.64	.912	.094	.043	.83	.84
2019	5,037	817.14	.945	.076	.033	.84	.85
2020	5,890	1,191.76	.931	.086	.037	.85	.85
Excluding “Feel others’ emotions” (*df* = 20)							
2008	6,773	729.39	.946	.072	.033	.80	.80
2009	5,619	560.39	.956	.069	.030	.81	.82
2010	1,375	214.25	.919	.084	.043	.78	.78
2011	5,292	358.97	.970	.057	.025	.81	.82
2012	1,476	140.99	.955	.064	.030	.79	.79
2013	5,139	530.72	.954	.070	.030	.81	.82
2014	6,511	577.63	.961	.065	.028	.82	.82
2015	506	43.01 [b]	.979	.048	.028	.82	.83
2017	6,062	755.35	.946	.078	.034	.82	.83
2018	807	126.12	.939	.081	.037	.82	.82
2019	5,037	388.10	.969	.060	.025	.83	.83
2020	5,890	622.09	.956	.071	.030	.83	.83

Wave	*n*	χ^2 [a]	CFI	RMSEA	SRMR	α	ω
Excluding "Am not interested in other people's problems" (*df* = 14)							
2008	6,773	365.84	.966	.061	.026	.77	.77
2009	5,619	264.37	.975	.056	.023	.79	.79
2010	1,375	96.80	.957	.066	.031	.75	.75
2011	5,292	188.46	.981	.049	.020	.79	.80
2012	1,476	68.01	.975	.051	.024	.76	.77
2013	5,139	235.57	.975	.055	.023	.79	.80
2014	6,511	347.13	.972	.060	.024	.80	.80
2015	506	27.84 [c]	.985	.044	.025	.80	.81
2017	6,062	366.68	.967	.064	.026	.80	.80
2018	807	56.87	.970	.062	.029	.80	.80
2019	5,037	204.06	.980	.052	.020	.81	.81
2020	5,890	263.79	.978	.055	.022	.81	.81
Excluding "Am interested in people" (*df* = 9)							
2008	6,769	271.16	.964	.066	.027	.72	.73
2009	5,619	199.47	.972	.061	.024	.74	.75
2010	1,374	69.57	.956	.070	.032	.71	.70
2011	5,287	129.68	.981	.050	.020	.75	.75
2012	1,475	57.81	.968	.061	.026	.72	.72
2013	5,135	156.90	.976	.057	.022	.75	.75
2014	6,507	243.97	.971	.063	.024	.75	.76
2015	505	20.48 [c]	.982	.050	.025	.76	.76
2017	6,061	264.96	.966	.069	.026	.75	.76
2018	806	49.58	.960	.075	.032	.75	.76
2019	5,034	131.45	.982	.052	.019	.77	.77
2020	5,884	171.27	.980	.055	.021	.77	.77

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation;

SRMR = standardized root mean square residual.

[a] All *p*s < .001 (if not indicated otherwise). [b] $p = .002$. [c] $p = .015$.

Openness

Unsatisfactory psychometric properties arose when Openness was measured with 10 items (see Table S14). In the majority of waves, the highest MI indicated a residual correlation between "Have difficulty understanding abstract ideas" and "Am not interested in abstract ideas," arguably due to their similar wording and overlapping content. For most waves, excluding "Have difficulty understanding abstract ideas" yielded stronger improvement in the fit indices than excluding "Have difficulty understanding abstract ideas," whereas the internal consistencies were not markedly different between the two options. Therefore, "Have difficulty understanding abstract ideas" was excluded.

As Table S14 shows, after the item was excluded, the model fit improved, but at the same time, the results indicated that another item needed to be removed. Now, a residual correlation between "Have a rich vocabulary" and "Use difficult words" was indicated by the highest MI in the majority of waves, not surprisingly due to the similarity of the two items. Consistently, lower factor loadings were observed for "Use difficult words" than for "Have a rich vocabulary." At the same time, the internal consistencies were higher without "Use difficult words" than without "Have a rich vocabulary." Therefore, we decided to exclude "Use difficult words."

Thereafter, the fit improved (see Table S14), and among the highest MIs, there was a residual correlation between "Have excellent ideas" and "Am full of ideas," reflecting the similar item content. In each wave, "Am full of ideas" exhibited a higher factor loading than "Have excellent ideas." Further, the internal consistencies were higher when "Have excellent ideas" was excluded instead of "Am full of ideas." Lastly, "Have excellent ideas" seemed somewhat evaluative (potentially also capturing interindividual differences in self-evaluation). Hence, "Have excellent ideas" was excluded.

As Table S14 shows, excluding this item improved the model fit. A residual correlation between “Have a vivid imagination” and “Am full of ideas” was among the highest MIs in each wave. Consistently, “Am full of ideas” had a higher factor loading than “Have a vivid imagination.” For the model fit and internal consistencies, excluding “Have a vivid imagination” yielded more favorable results than excluding “Am full of ideas.” Further, “Have a vivid imagination” might be considered the opposite of “Do not have a good imagination,” indicating a doubling of the items. Therefore, “Have a vivid imagination” was excluded, resulting in overall acceptable psychometric criteria (see Table S14).

Table S14

Psychometric Criteria for the Item Selection for Openness

Wave	*n*	$\chi^{2\,a}$	CFI	RMSEA	SRMR	α	ω
All items (*df* = 35)							
2008	6,773	3,825.72	.736	.126	.067	.77	.77
2009	5,619	3,328.19	.729	.129	.069	.77	.77
2010	1,375	741.79	.712	.121	.069	.74	.73
2011	5,292	3,066.78	.736	.128	.068	.77	.77
2012	1,476	782.58	.748	.120	.066	.76	.76
2013	5,139	2,797.07	.741	.124	.066	.76	.76
2014	6,512	3,582.73	.739	.125	.067	.76	.76
2015	506	324.22	.756	.128	.068	.78	.78
2017	6,062	3,329.63	.747	.125	.066	.77	.77
2018	807	506.17	.742	.129	.070	.77	.77
2019	5,037	2,826.80	.747	.126	.065	.77	.77
2020	5,890	3,044.55	.760	.121	.064	.77	.77
Excluding “Have difficulty understanding abstract ideas” (*df* = 27)							
2008	6,769	2,133.37	.808	.107	.057	.73	.73
2009	5,619	1,851.55	.807	.110	.059	.74	.73
2010	1,374	403.10	.803	.101	.057	.71	.71
2011	5,287	1,690.85	.813	.108	.057	.74	.73
2012	1,475	542.22	.780	.114	.061	.73	.73
2013	5,135	1,550.53	.814	.105	.055	.73	.73
2014	6,507	1,987.94	.814	.106	.058	.73	.72
2015	505	227.29	.782	.121	.063	.75	.74
2017	6,061	1,939.30	.812	.108	.057	.74	.74
2018	806	243.72	.844	.100	.055	.75	.74
2019	5,034	1,583.29	.814	.107	.055	.74	.73
2020	5,884	1,841.60	.816	.107	.056	.74	.74
Excluding “Use difficult words” (*df* = 20)							
2008	6,769	1,243.62	.869	.095	.048	.72	.72
2009	5,619	1,069.56	.871	.097	.049	.73	.72
2010	1,374	214.89	.876	.084	.046	.69	.69
2011	5,287	991.75	.873	.096	.048	.73	.73
2012	1,475	330.27	.841	.103	.053	.71	.71
2013	5,135	807.06	.886	.088	.044	.72	.72
2014	6,507	1,093.74	.881	.091	.047	.72	.71
2015	505	156.16	.829	.116	.057	.74	.73
2017	6,061	1,170.77	.867	.097	.048	.72	.72
2018	806	136.96	.900	.085	.046	.73	.73
2019	5,034	1,027.14	.861	.100	.049	.73	.72
2020	5,884	1,080.49	.874	.095	.047	.73	.73

Wave	*n*	χ^2 [a]	CFI	RMSEA	SRMR	α	ω
Excluding "Have excellent ideas" (*df* = 14)							
2008	6,769	739.75	.877	.088	.040	.67	.68
2009	5,619	577.22	.892	.085	.039	.69	.69
2010	1,374	132.62	.882	.079	.039	.65	.65
2011	5,287	590.51	.883	.088	.040	.69	.69
2012	1,475	220.10	.842	.100	.049	.67	.67
2013	5,135	550.00	.882	.086	.040	.67	.68
2014	6,507	637.13	.886	.083	.039	.67	.67
2015	505	113.12	.825	.118	.054	.69	.69
2017	6,061	732.22	.870	.092	.042	.68	.68
2018	806	90.11	.900	.082	.042	.69	.69
2019	5,034	592.62	.875	.091	.041	.68	.68
2020	5,884	702.93	.873	.091	.042	.68	.69
Excluding "Have a vivid imagination" (*df* = 9)							
2008	6,769	156.04	.965	.049	.022	.63	.64
2009	5,619	147.12	.962	.052	.023	.65	.65
2010	1,374	44.08	.951	.053	.026	.60	.60
2011	5,287	123.87	.967	.049	.022	.65	.65
2012	1,475	46.57	.960	.053	.027	.64	.64
2013	5,135	122.76	.964	.050	.023	.63	.63
2014	6,507	184.15	.955	.055	.025	.63	.63
2015	505	28.76 [b]	.953	.066	.033	.67	.67
2017	6,060	205.53	.950	.060	.027	.64	.64
2018	805	34.03	.952	.059	.032	.64	.64
2019	5,034	143.40	.959	.054	.025	.64	.64
2020	5,884	185.15	.954	.058	.026	.64	.64

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation;

SRMR = standardized root mean square residual.

[a] All *p*s < .001 (if not indicated otherwise). [b] $p = .001$.

Psychometric Analyses

Table S15

Wave-Specific Psychometric Properties in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel

Parameter	2008	2009	2010	2011	2012	2013	2014	2015	2017	2018	2019	2020
n	5,445	5,455	1,042	5,199	1,251	5,079	5,880	422	5,527	675	4,969	4,748
Neuroticism												
α	.78	.79	.75	.79	.78	.80	.81	.85	.80	.81	.82	.81
ω	.81	.81	.79	.84	.82	.85	.83	.87	.85	.83	.84	.84
Extraversion												
α	.79	.80	.77	.80	.81	.81	.81	.79	.82	.81	.82	.82
ω	.83	.81	.80	.83	.85	.85	.85	.84	.85	.85	.83	.84
Conscientiousness												
α	.68	.71	.70	.70	.69	.70	.70	.73	.70	.68	.70	.70
ω	.74	.73	.73	.73	.72	.74	.74	.78	.73	.71	.76	.72
Agreeableness												
α	.73	.75	.71	.75	.72	.75	.76	.77	.76	.76	.77	.78
ω	.76	.77	.73	.77	.75	.78	.78	.79	.79	.81	.79	.80
Openness												
α	.64	.66	.62	.66	.65	.64	.65	.69	.65	.63	.65	.65
ω	.67	.68	.68	.68	.67	.67	.68	.74	.68	.67	.69	.71

Measurement Invariance

Table S16

Fit Indices for Testing for Measurement Invariance (MI) for the Big Five in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel Based on Five Waves

Model	*n*	CFI	RMSEA	SRMR
Neuroticism	7,195			
Baseline		.967	.031	.031
Metric MI		.965	.031	.038
Scalar MI		.954	.034	.041
Extraversion	7,195			
Baseline		.974	.029	.030
Metric MI		.972	.030	.039
Scalar MI		.960	.034	.043
Conscientiousness	7,195			
Baseline		.962	.029	.034
Metric MI		.959	.029	.039
Scalar MI		.943	.033	.045
Agreeableness	7,195			
Baseline		.961	.030	.028
Metric MI		.958	.030	.034
Scalar MI		.946	.033	.039
Openness	7,195			
Baseline		.961	.029	.035
Metric MI		.959	.028	.038
Scalar MI		.950	.030	.041

Note. Personality assessments were included for 2008, 2011, 2014, 2017, and 2020. For each trait, focal points (representing birth years) ranged from 1933 to 2000. Metric MI requires the same factor loadings across both years of measurement and years of birth. Scalar MI additionally requires the same item intercepts across both years of measurement and years of birth.

CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Fixed Effects Modeling

Table S17

Parameter Estimates for the Effect of the First Wave on Personality in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel

Model	Estimate	*SE*	*t*	*p*
	Neuroticism			
Mean score	0.534	0.076	6.983	< .001
Contributing to the mean score				
Seldom feel blue (recoded)	0.325	0.111	2.927	.003
Am relaxed most of the time (recoded)	0.384	0.104	3.686	< .001
Get upset easily	0.194	0.097	2.008	.045
Get stressed out easily	0.285	0.087	3.280	.001
Get irritated easily	0.163	0.097	1.688	.091
Worry about things	0.945	0.102	9.293	< .001
Not contributing to the mean score				
Am easily disturbed	0.524	0.098	5.331	< .001
Change my mood a lot	0.168	0.098	1.722	.085
Have regular mood swings	0.189	0.096	1.978	.048
Often feel blue	0.179	0.092	1.949	.051
	Extraversion			
Mean score	0.267	0.068	3.948	< .001
Contributing to the mean score				
Am quiet around strangers (recoded)	0.036	0.091	0.396	.692
Don’t talk a lot (recoded)	0.099	0.093	1.064	.288
Feel comfortable around people	0.227	0.095	2.399	.016
Don’t mind being the center of attention	0.066	0.098	0.679	.497
Start conversations	0.175	0.088	1.987	.047
Liven things up	0.605	0.089	6.767	< .001
Not contributing to the mean score				
Keep in the background (recoded)	0.153	0.093	1.638	.101
Have little to say (recoded)	0.314	0.106	2.969	.003
Talk to a lot of different people at parties	0.064	0.087	0.734	.463
Don’t like to draw attention to myself (recoded)	−0.178	0.101	−1.761	.078

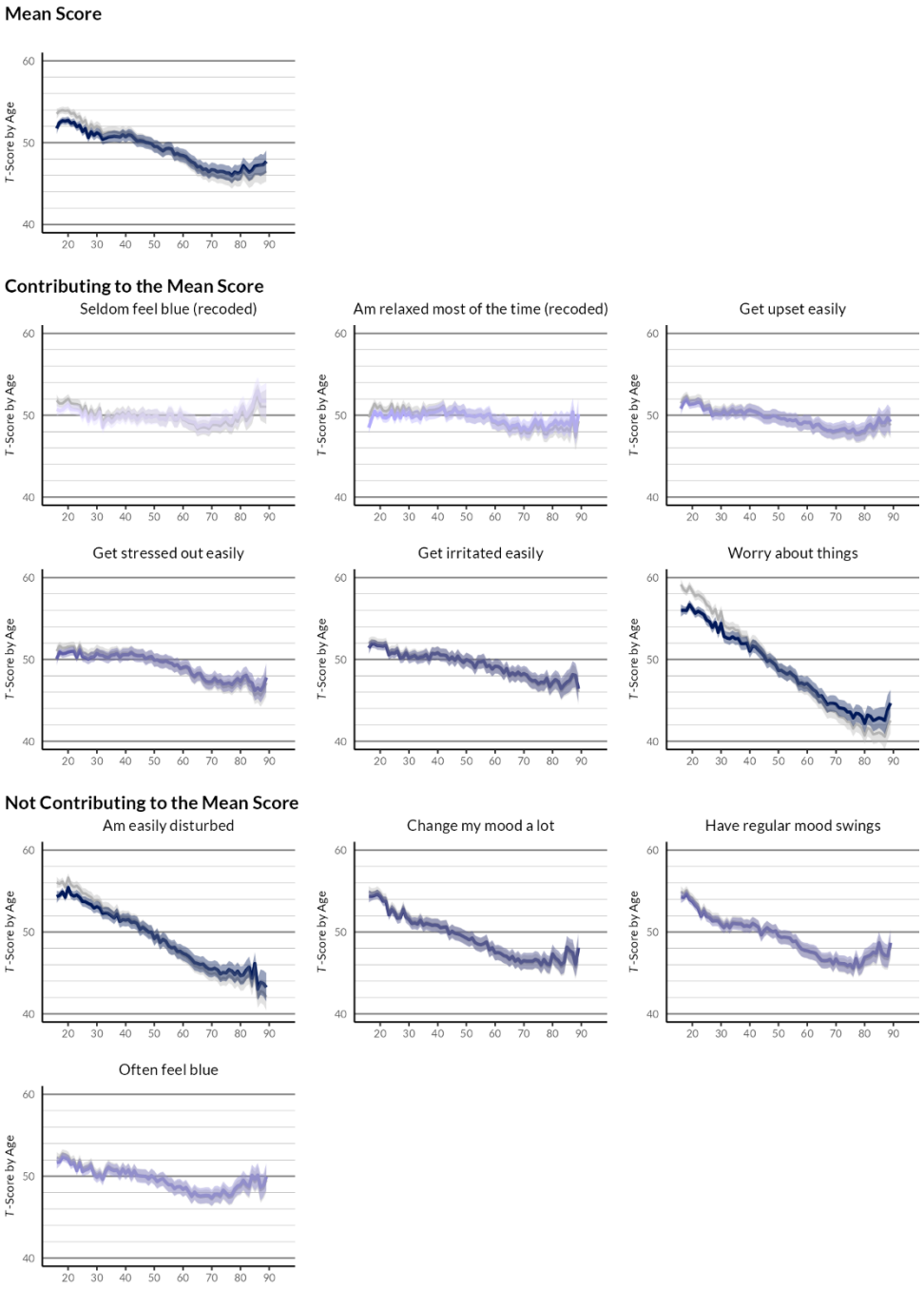
Model	Estimate	*SE*	*t*	*p*
	Conscientiousness			
Mean score	0.552	0.079	6.997	< .001
Contributing to the mean score				
Pay attention to details	0.446	0.105	4.253	< .001
Get chores done right away	0.292	0.095	3.064	.002
Like order	0.340	0.096	3.532	< .001
Often forget to put things back in their proper place (recoded)	0.110	0.096	1.148	.251
Follow a schedule	0.348	0.101	3.452	.001
Shirk my duties (recoded)	0.640	0.103	6.226	< .001
Not contributing to the mean score				
Am always prepared	0.296	0.109	2.705	.007
Leave my belongings around (recoded)	0.333	0.091	3.673	< .001
Make a mess of things (recoded)	0.284	0.098	2.905	.004
Am exacting in my work	0.752	0.099	7.583	< .001
	Agreeableness			
Mean score	0.665	0.084	7.892	< .001
Contributing to the mean score				
Am not really interested in others (recoded)	0.271	0.107	2.525	.012
Take time out for others	0.614	0.108	5.673	< .001
Sympathize with others' feelings	0.840	0.102	8.211	< .001
Feel little concern for others (recoded)	0.100	0.118	0.852	.394
Make people feel at ease	0.677	0.103	6.558	< .001
Insult people (recoded)	0.317	0.097	3.262	.001
Not contributing to the mean score				
Am interested in people	0.664	0.100	6.665	< .001
Am not interested in other people's problems (recoded)	0.172	0.114	1.515	.130
Am sentimental	0.291	0.095	3.070	.002
Feel others' emotions	0.596	0.100	5.980	< .001

Model	Estimate	*SE*	*t*	*p*
	Openness			
Mean score	0.589	0.082	7.201	< .001
Contributing to the mean score				
Do not have a good imagination (recoded)	0.069	0.118	0.584	.559
Am full of ideas	0.567	0.092	6.190	< .001
Am not interested in abstract ideas (recoded)	0.056	0.107	0.522	.602
Am quick to understand things	0.391	0.098	4.002	< .001
Spend time reflecting on things	0.725	0.109	6.663	< .001
Have a rich vocabulary	0.420	0.089	4.747	< .001
Not contributing to the mean score				
Have difficulty understanding abstract ideas (recoded)	0.305	0.098	3.121	.002
Have a vivid imagination	0.407	0.089	4.557	< .001
Have excellent ideas	0.405	0.102	3.974	< .001
Use difficult words	0.066	0.088	0.748	.454

Note. The first wave was dummy-coded as 1 = first wave and 0 = all other waves. Parameters were estimated in separate fixed effects models with dummies for age. Panel-robust standard errors (as proposed by Arellano, 1987) were computed to account for serial correlation and heteroscedasticity.

Figure S11

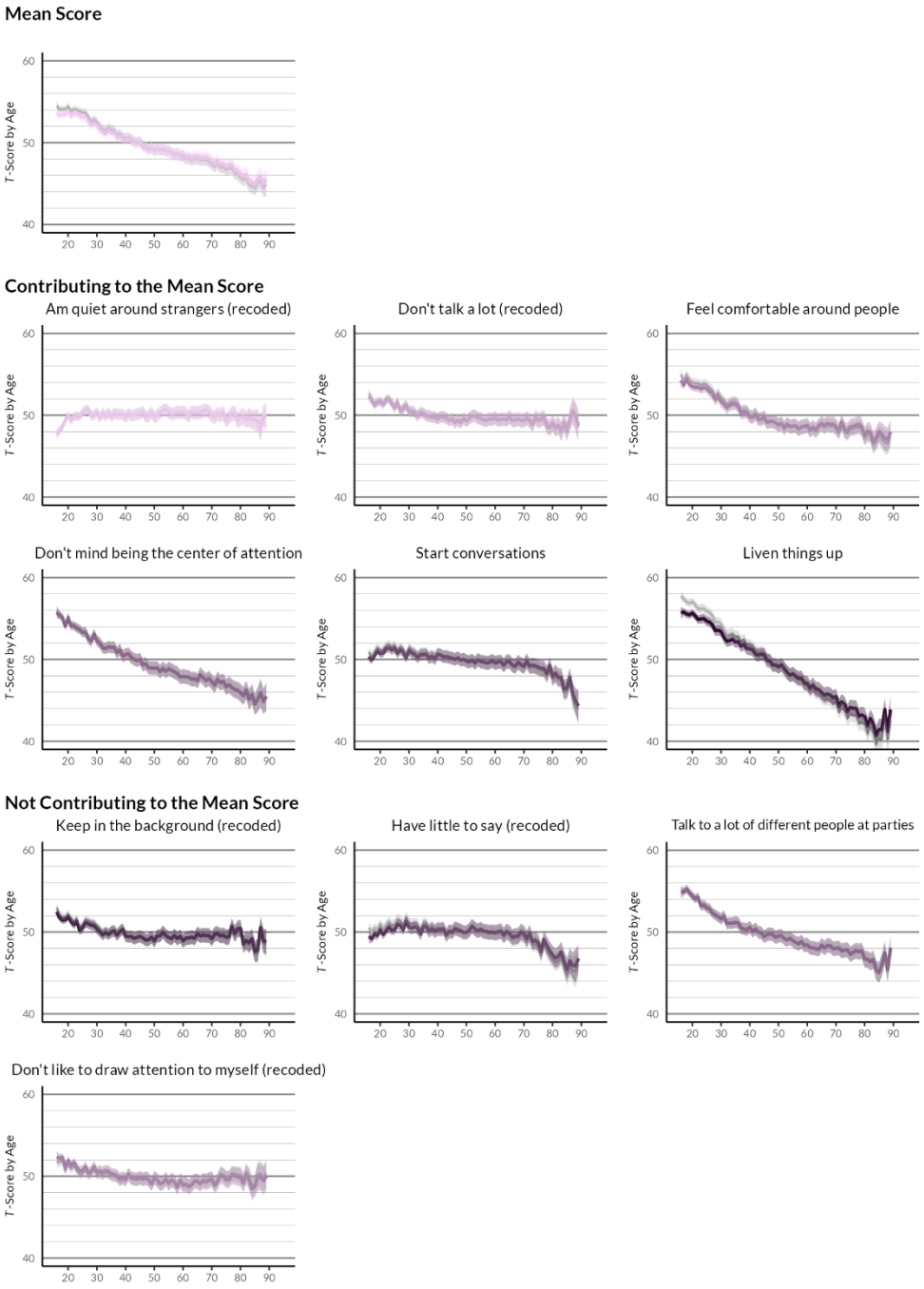
Within-Person Mean-Level Age Trajectories for Neuroticism in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S12

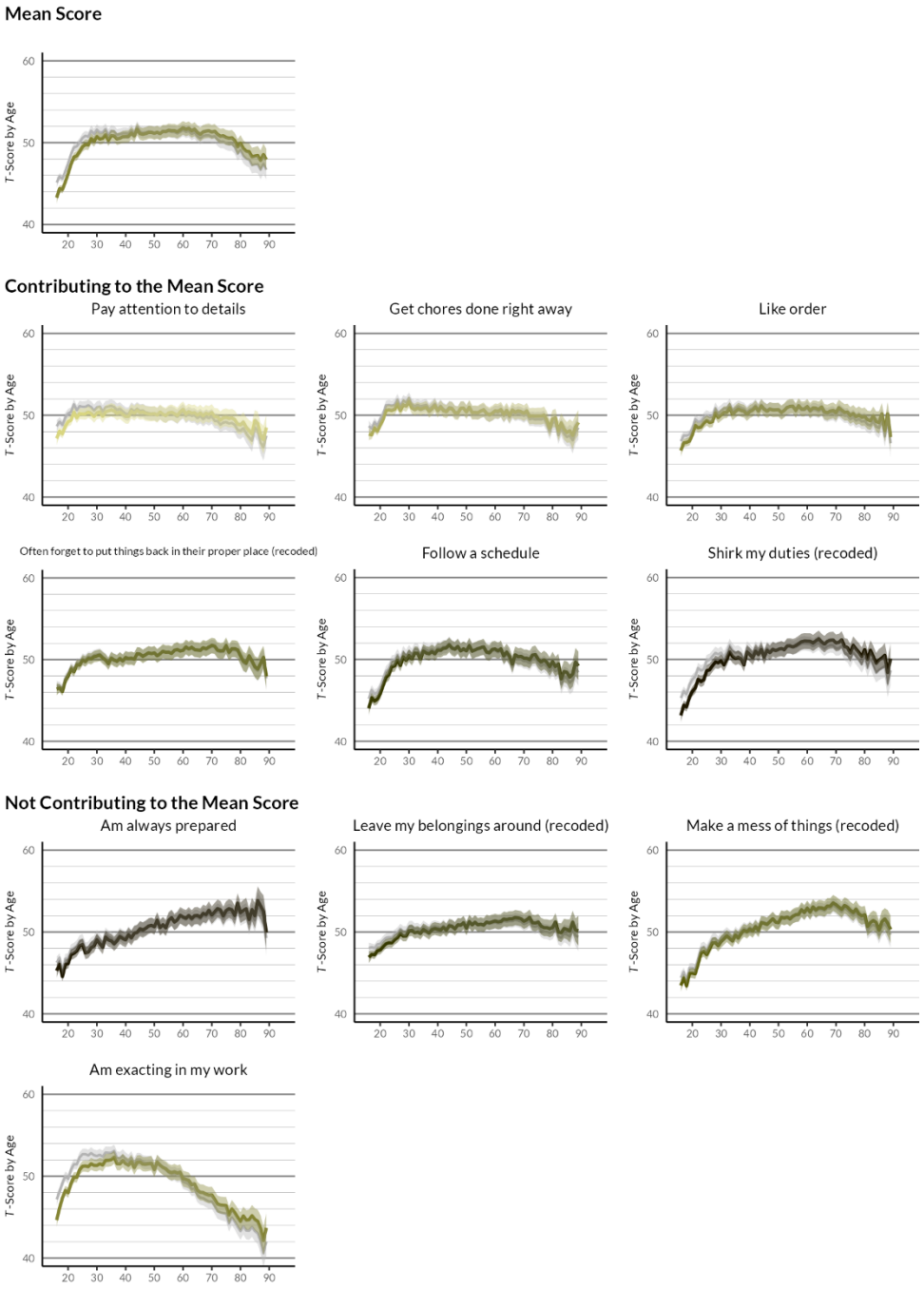
Within-Person Mean-Level Age Trajectories for Extraversion in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S13

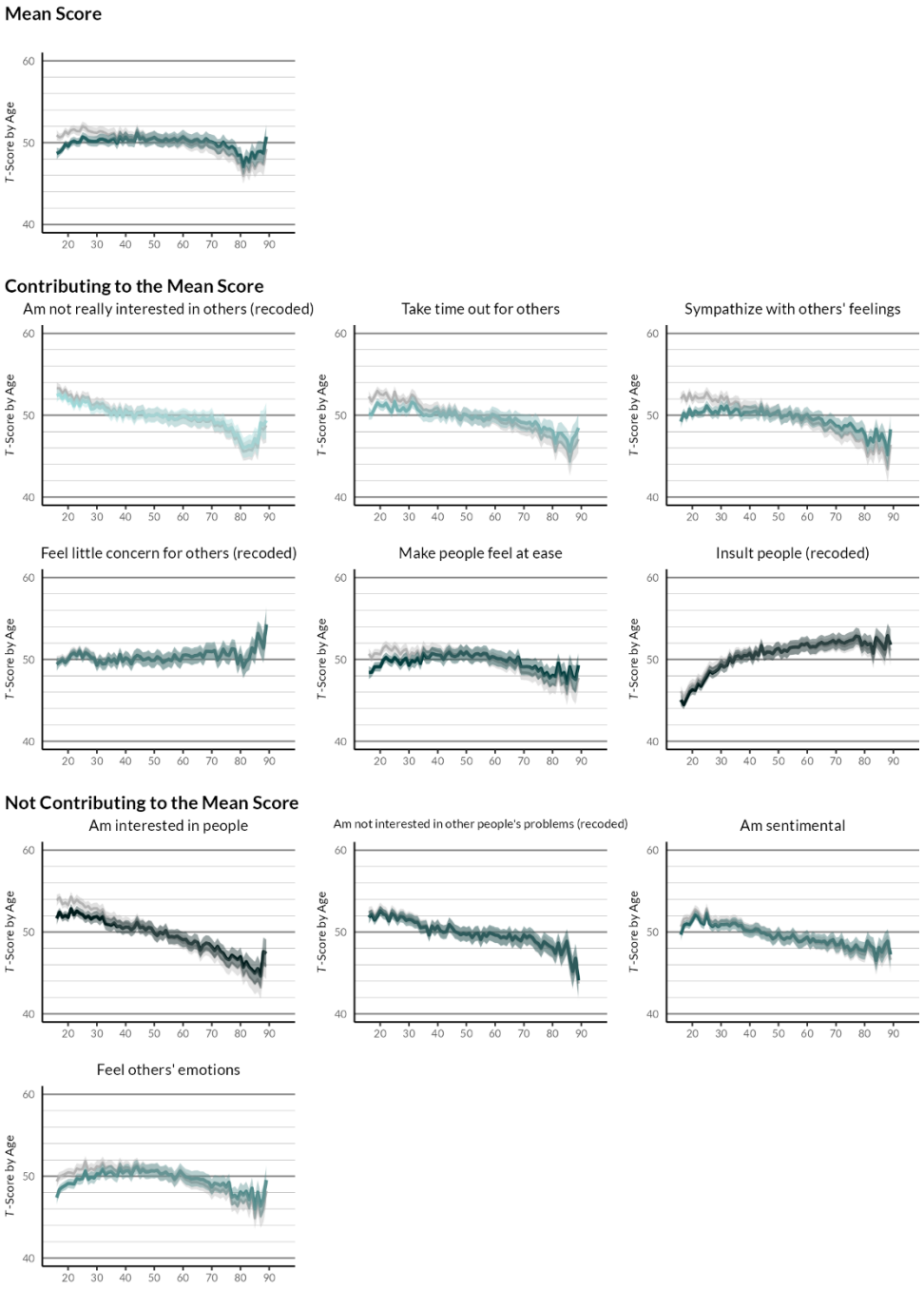
Within-Person Mean-Level Age Trajectories for Conscientiousness in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S14

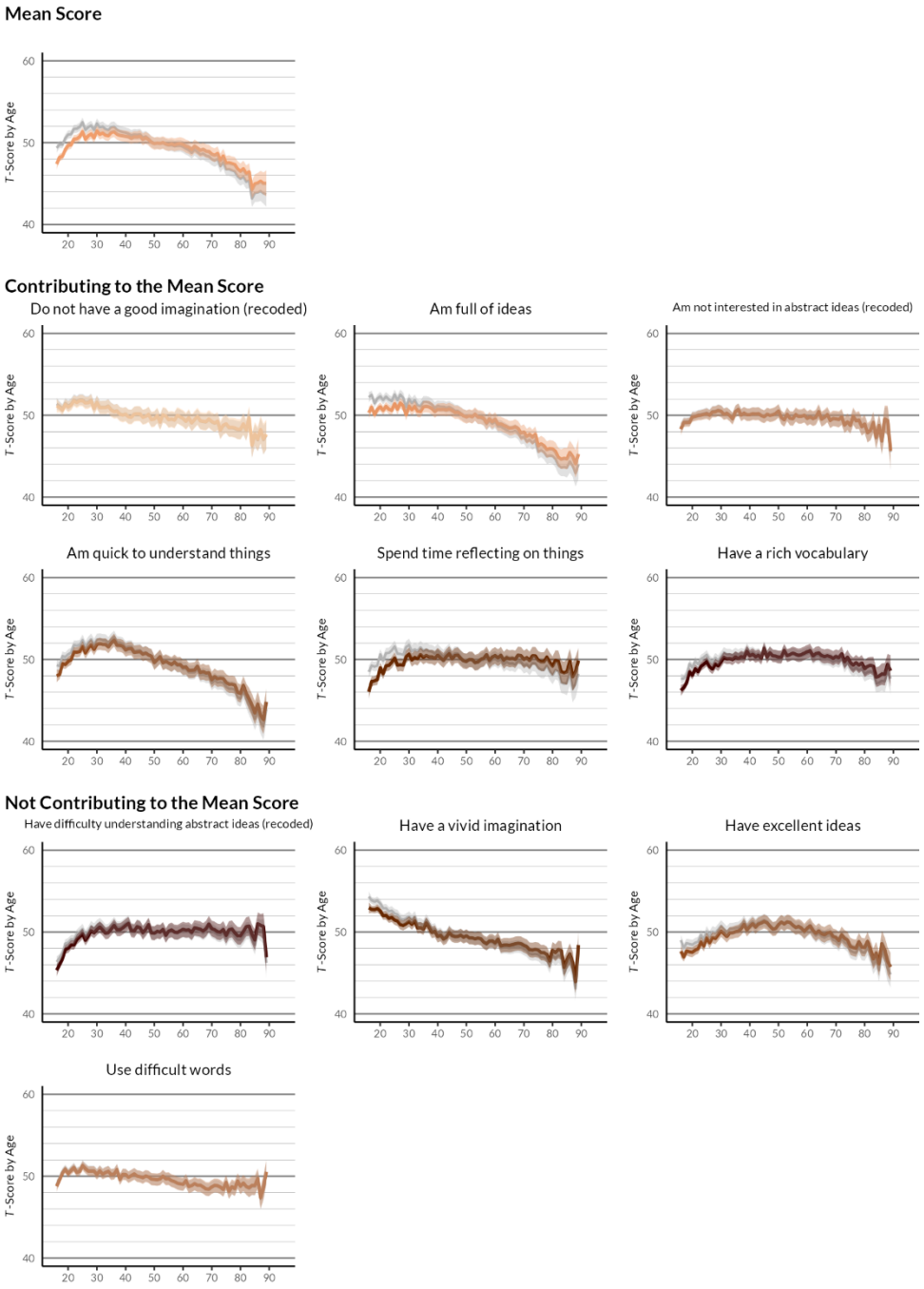
Within-Person Mean-Level Age Trajectories for Agreeableness in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Figure S15

Within-Person Mean-Level Age Trajectories for Openness in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the panel-robust standard error.

Additional Analyses

Comparing Fixed Effects Modeling, Standard Multilevel Modeling, and Between-Person Differences

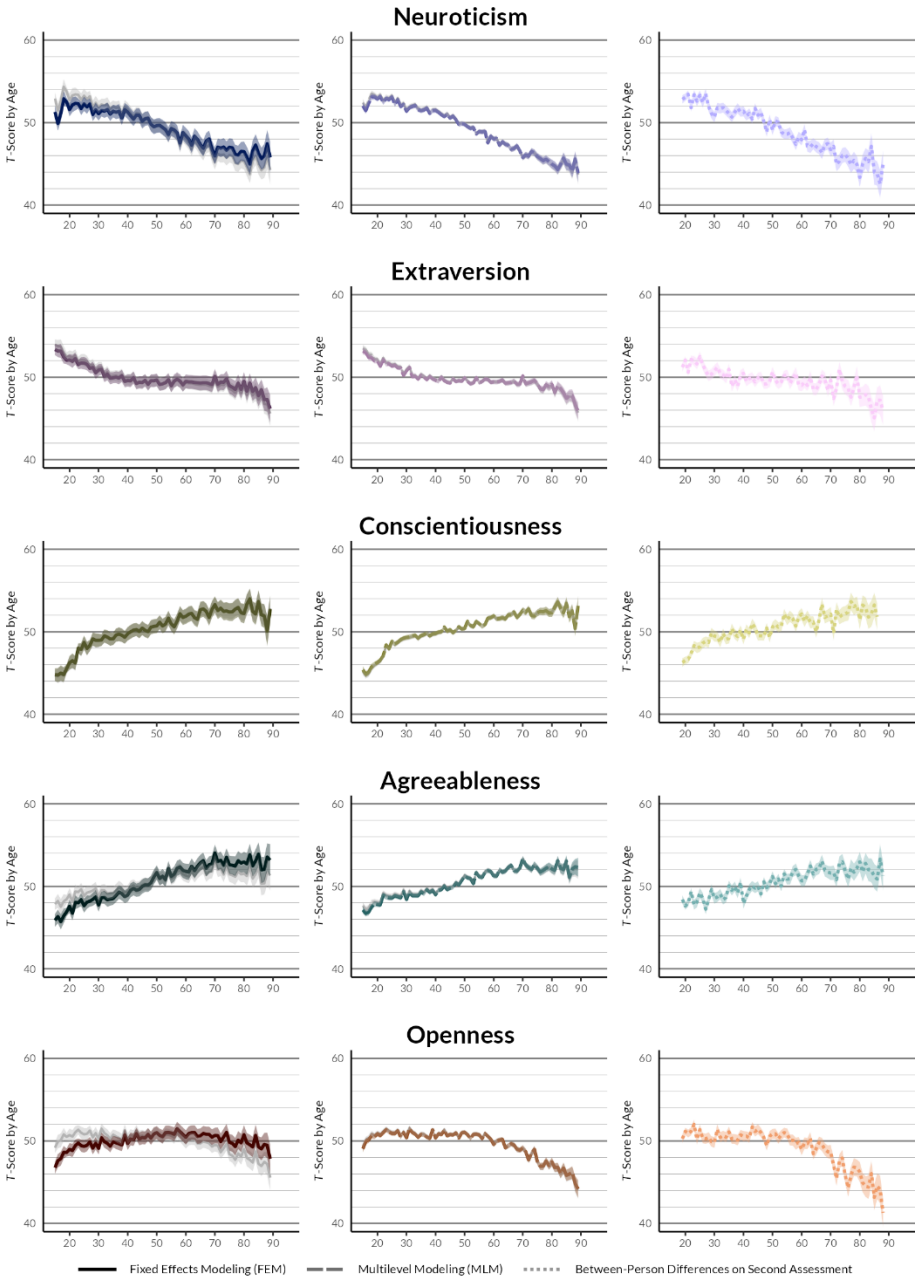
With our fixed effects modeling (FEM) approach, we considered only within-person information to estimate personality age trajectories. As explained in more detail in the main text, we also reanalyzed the data with conventional multilevel modeling (MLM) to obtain estimates that blended within-person and between-person effects. In these supplemental analyses, we now report complementary results of cross-sectional analyses that use only between-person information.

For these cross-sectional analyses, we relied on the second measurement occasion, as the first may have been distorted by an initial elevation bias (e.g., Shrout et al., 2018). Similar to the analyses with FEM and MLM, we added age to the regression models as a dummy-coded predictor variable; and in Study 2, we controlled for differences in mode. Parameters were obtained in a linear regression framework via ordinary least squares (OLS) estimation.

The trajectories are shown in Figure S16 for Study 1, in Figure S17 for Study 2, and in Figure S18 for Study 3. Results for a particular year of age were depicted in the trajectories when at least $n = 30$ participants provided data for that age. For Conscientiousness in Study 1, the age span was restricted to avoid a gap in the age range. Accordingly, for Study 1, the age range plotted for each trait was 19 to 88 (except for Conscientiousness, which had a range of 19 to 86). For Study 2, the range for each trait was 18 to 88; and for Study 3, it was 17 to 81.

Figure S16

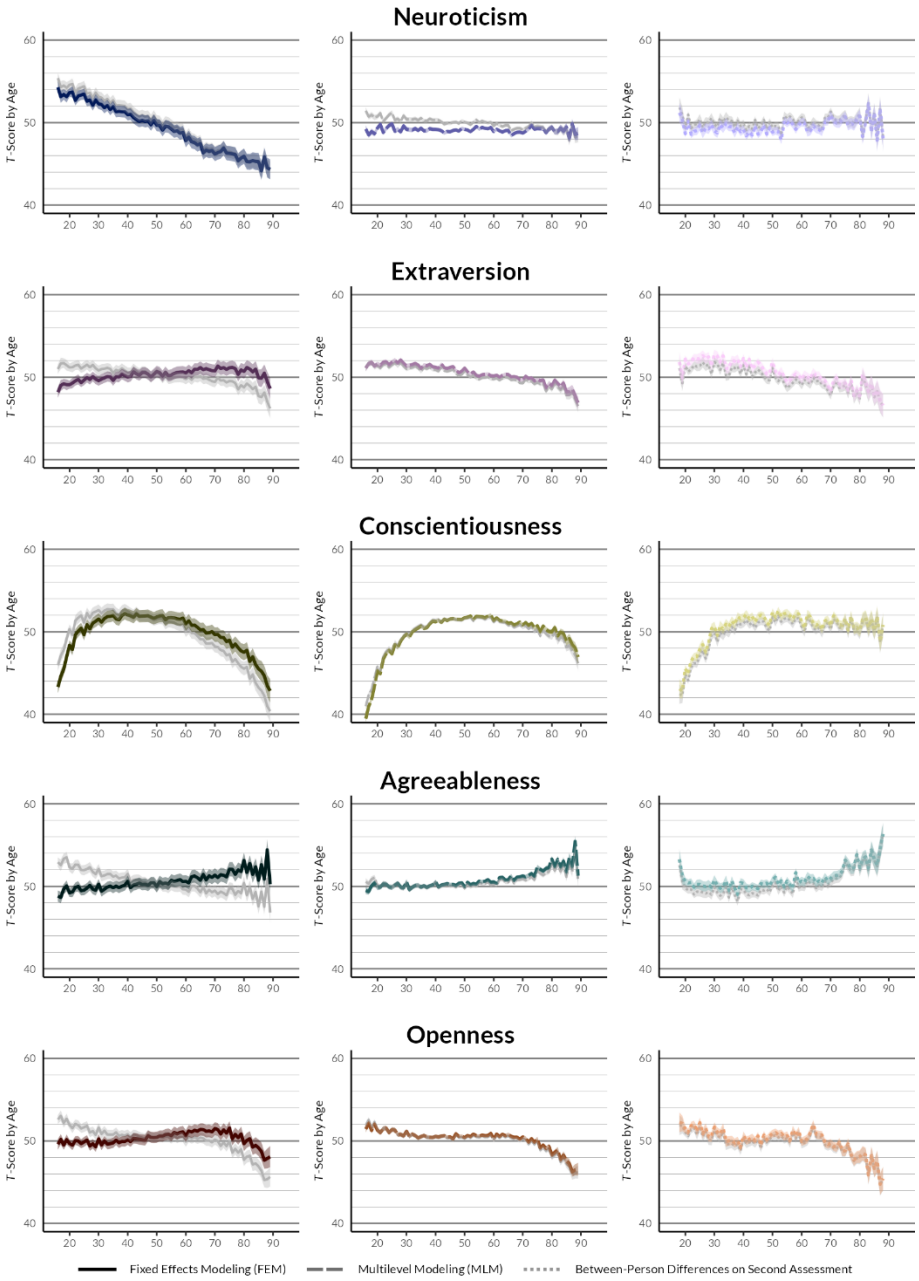
Mean-Level Age Trajectories for the Big Five Scores Based on Fixed Effects Modeling (FEM; Left Panel), Multilevel Modeling (MLM; Middle Panel), and Between-Person Differences in the Household, Income and Labour Dynamics in Australia (HILDA) Survey (Study 1)



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the standard error.

Figure S17

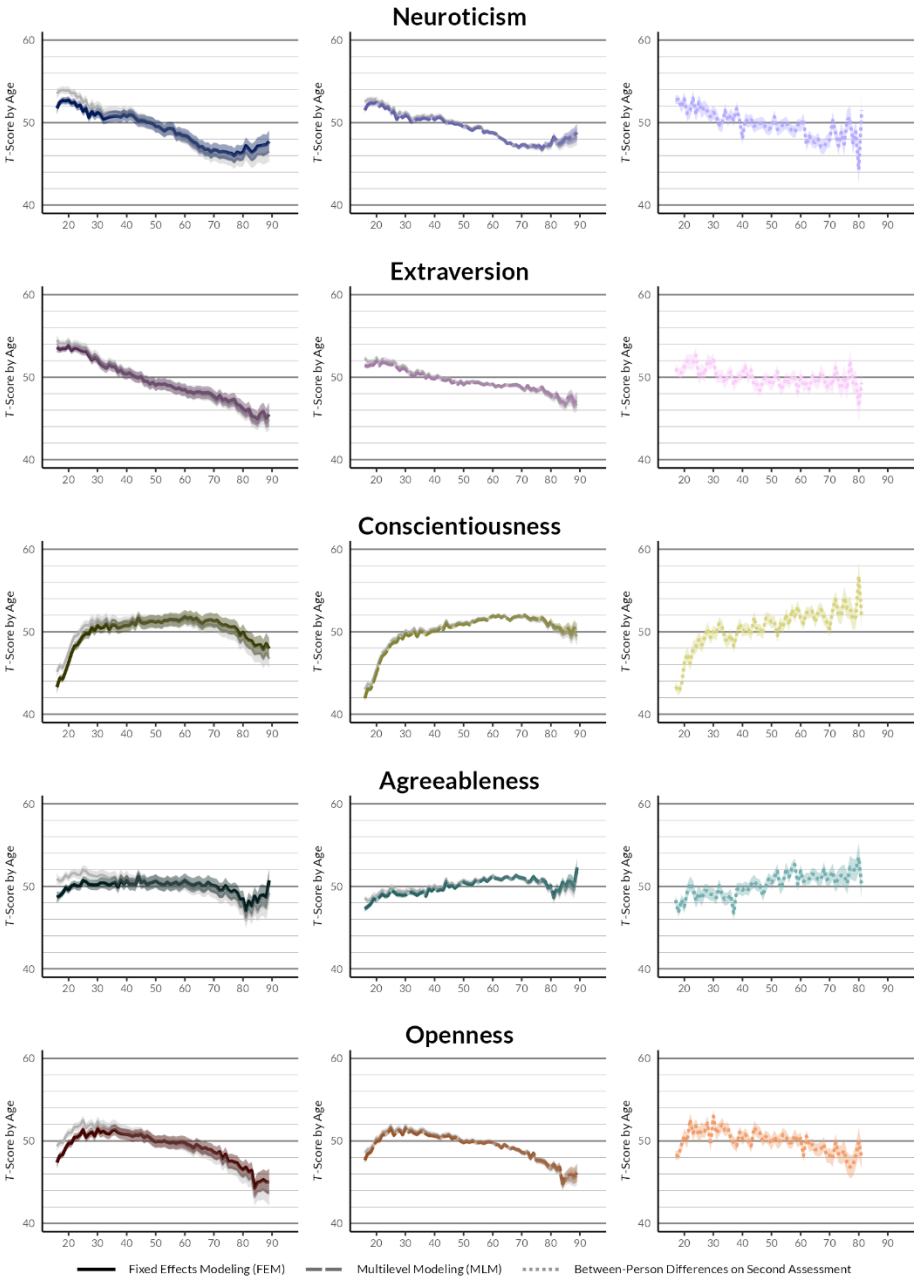
Mean-Level Age Trajectories for the Big Five Scores Based on Fixed Effects Modeling (FEM; Left Panel), Multilevel Modeling (MLM; Middle Panel), and Between-Person Differences in the Socio-Economic Panel (SOEP; Study 2)



Note. The grey line indicates the trajectory when the first wave and mode were not controlled for. The shaded area represents the standard error.

Figure S18

Mean-Level Age Trajectories for the Big Five Scores Based on Fixed Effects Modeling (FEM; Left Panel), Multilevel Modeling (MLM; Middle Panel), and Between-Person Differences in the Longitudinal Internet Studies for the Social Sciences (LISS) Panel (Study 3)



Note. The grey line indicates the trajectory when the first wave was not controlled for. The shaded area represents the standard error.

Additionally Including Participants With Only One Personality Assessment in Standard Multilevel Modeling

We used FEM to estimate personality age trajectories that relied on only within-person changes. As a necessary prerequisite for these longitudinal analyses, participants were included only if they provided personality data for at least two time points. By contrast, MLM without person-mean centering has the potential advantage that participants with only one measurement can be included in the analyses (but at the expense of introducing between-person information in the estimates). To examine the potential impact of excluding participants with a single assessment, we next compared the results of an MLM analysis that was based on all available participants with the MLM analysis reported in the previous section, which included only those participants who provided data for at least two measurement occasions. Thus, for each study, we extended the data that we had previously analyzed by also including participants who had only one personality assessment. The resulting sample sizes are reported in Table S18.

Personality scores were transformed into *T*-scores by using the participants' first assessment (note that this is somewhat different from the other analyses where we used the *second* assessment for standardization). To estimate age trajectories, we used MLM with the same model that we used to compare the age trends from FEM with MLM as reported in the previous section: We included age in the regression models as a categorical variable. We limited the estimation to years of age for which we had at least $n = 30$ observations. We added the same control variables to the models (i.e., dummy-coded first wave [Studies 1 to 3] and response mode [Study 2]), without random slopes for individuals. The modeled age range was the same as in the previous analyses with FEM and MLM (i.e., for Study 1, 15 to 91; for Study 2, 16 to 93;

for Study 3, 16 to 89). Again, for a consistent display of results across studies, age trajectories were plotted up to age 89.

The resulting age trajectories are presented in Figure S19, which additionally contains MLM-based age trajectories for the same model when only participants with at least two assessments were included (see the middle panel in Figures S16 to S18, as well as Figure 4 in the manuscript). As a general result, additionally including participants *without* longitudinal data in the analyses with MLM provided virtually the same age trends as when only participants *with* longitudinal data were included. Consequently, we found no evidence that excluding participants who provided only a single personality assessment changed the results in any meaningful way.

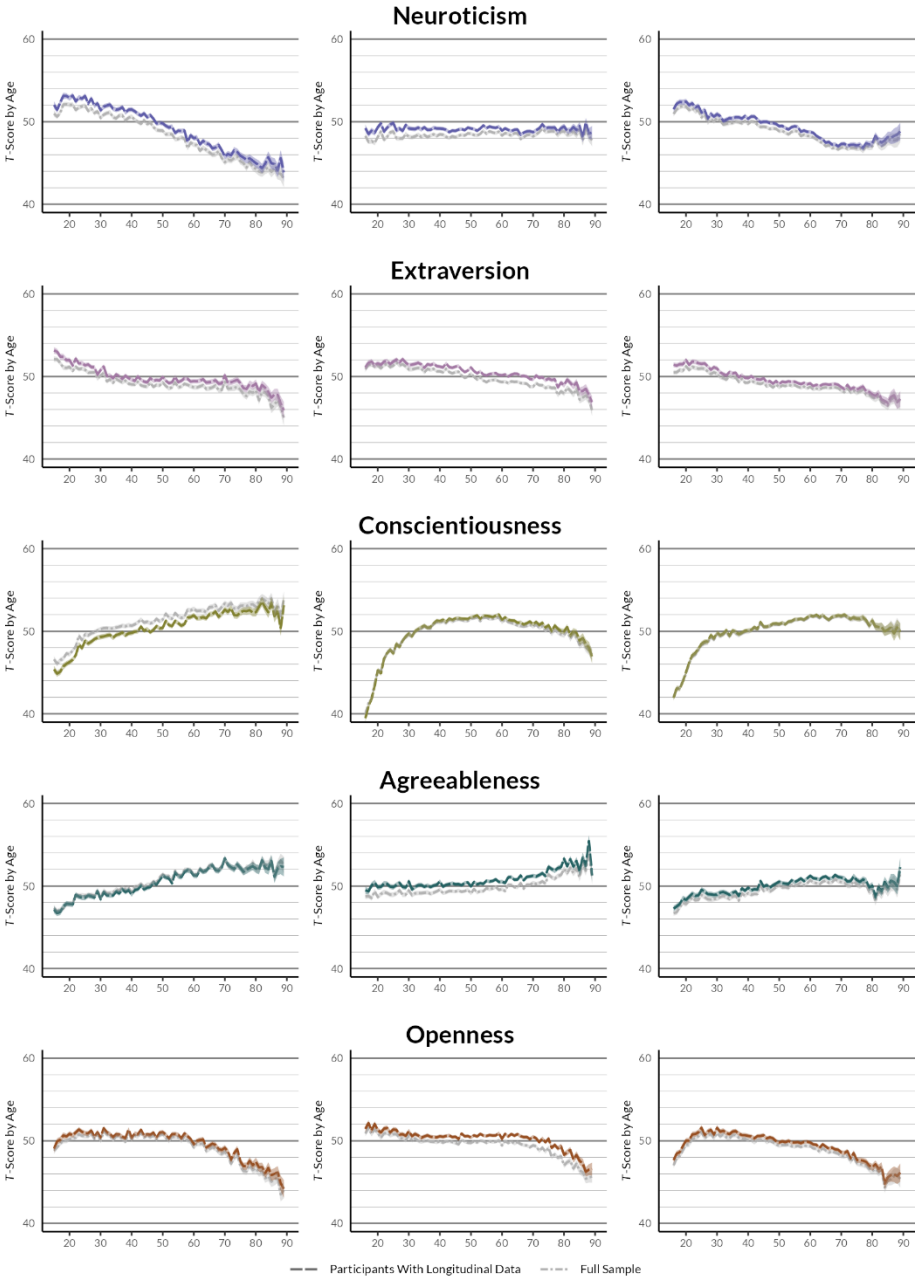
Table S18

Sample Sizes and Observations by Trait and Study When Including Participants With Only One Personality Assessment

Trait	*n*	Observations
Study 1: Household, Income and Labour Dynamics in Australia (HILDA) Survey		
Neuroticism	22,230	52,571
Extraversion	22,172	52,242
Conscientiousness	22,225	52,495
Agreeableness	22,320	52,896
Openness	22,179	52,302
Study 2: Socio-Economic Panel (SOEP)		
Neuroticism	33,144	84,585
Extraversion	33,150	84,548
Conscientiousness	33,074	84,278
Agreeableness	33,168	84,552
Openness	33,078	84,070
Study 3: Longitudinal Internet Studies for the Social Sciences (LISS) Panel		
Neuroticism	14,734	50,263
Extraversion	14,734	50,263
Conscientiousness	14,734	50,263
Agreeableness	14,734	50,263
Openness	14,734	50,263

Figure S19

Mean-Level Age Trajectories for the Big Five Scores Based on Multilevel Modeling (MLM) by Different Inclusion Criteria in the Household, Income and Labour Dynamics in Australia (HILDA) Survey (Left Panel), the Socio-Economic Panel (SOEP; Middle Panel), and the Longitudinal Internet Studies for the Social Sciences (LISS) Panel (Right Panel)



Note. The shaded area represents the standard error.

Quantifying the Similarity of Item Trajectories Across Studies

To quantify the similarity of the item trends, we calculated the Fréchet distance (Fréchet, 1906; see also Alt & Godau, 1995; Genolini et al., 2016) as a measure of similarity between trajectories. As an analogy for the Fréchet distance, imagine a person who is walking their dog on a leash. The two are on separate paths (i.e., in our case, the trajectories of two different items), and each can move at different speeds but cannot go backwards. The shortest leash that is necessary for the walk represents the Fréchet distance (Alt & Godau, 1995). In other words, the Fréchet distance is the longest distance between two item trajectories when each point on the one trajectory (i.e., the item value for a specific year of age) is linked with the nearest point on the other trajectory while "traveling" only forward on each trajectory (Genolini et al., 2016). The smaller the Fréchet distance, the larger the similarity between the trajectories.

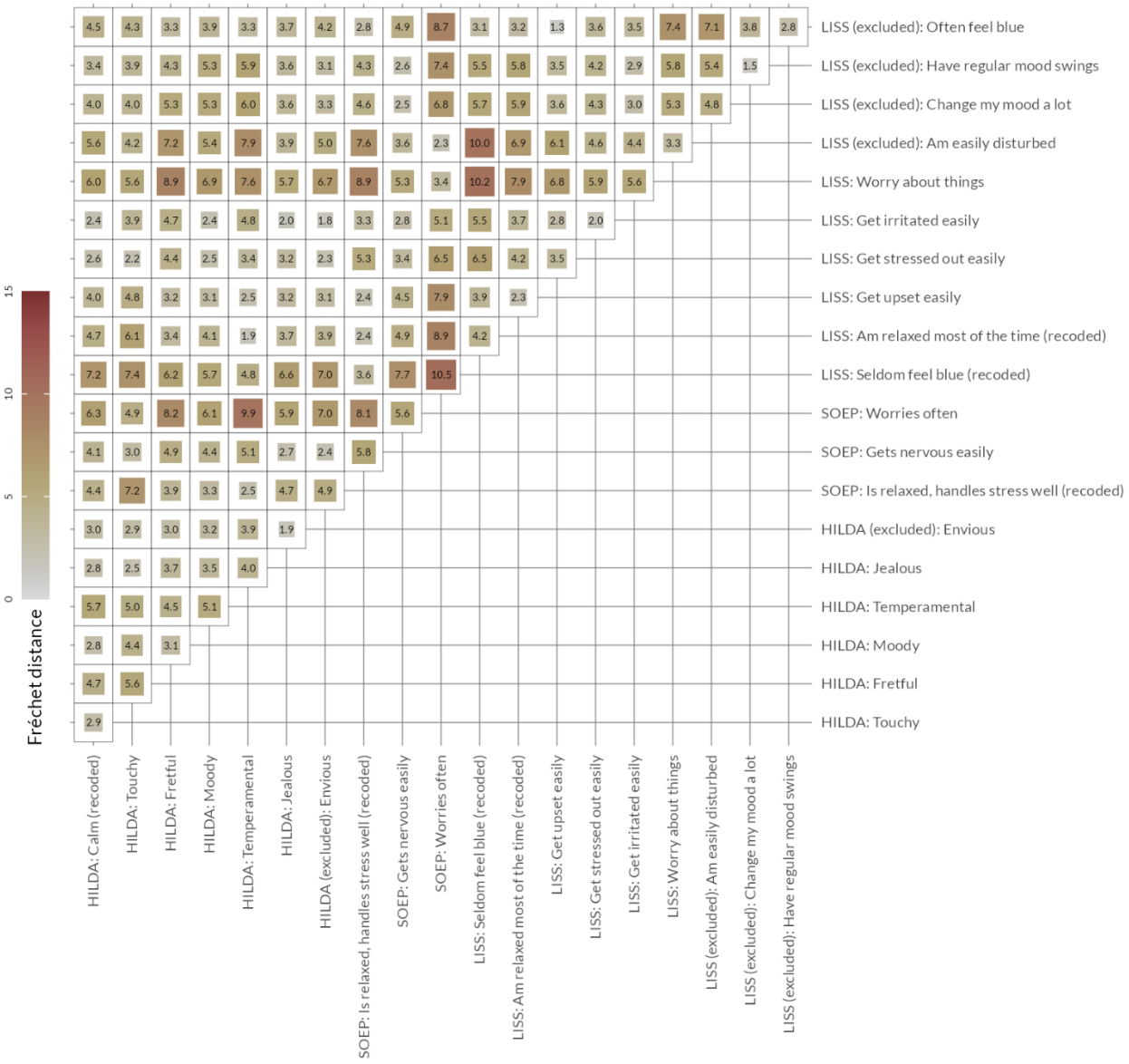
For the present analyses, we calculated the Fréchet distance for each pair of items that were used to measure the same personality trait. Thereby, it was also possible to estimate the similarity between trends for items from different studies. The Fréchet distance was applied to the trajectories from FEM where we controlled for the effects of the first wave (Studies 1 to 3) and the response mode (Study 2; see Figures S1 to S15). To obtain an identical age range across the item trajectories, we included estimates only for years 16 to 89. Figure S20 illustrates the results for Neuroticism, Figure S21 for Extraversion, Figure S22 for Conscientiousness, Figure S23 for Agreeableness, and Figure S24 for Openness.

Quantifying the similarity of the item trends by means of the Fréchet distance generally supported the visual inspection of the patterns: For some traits, items within a study showed relatively similar age trajectories (e.g., Extraversion items in Study 2 [SOEP]; see Figure S21), but for other traits, the items trajectories were less alike within a study (e.g., Agreeableness items

in Study 1 [HILDA]; see Figure S23). As another example, the Fréchet distance indeed indicated that developmental patterns for the Neuroticism items “Worries often” (Study 2 [SOEP]) and “Worry about things” (Study 3 [LISS]) were similar but at the same time distinct from the other item trends (see Figure S20).

Figure S20

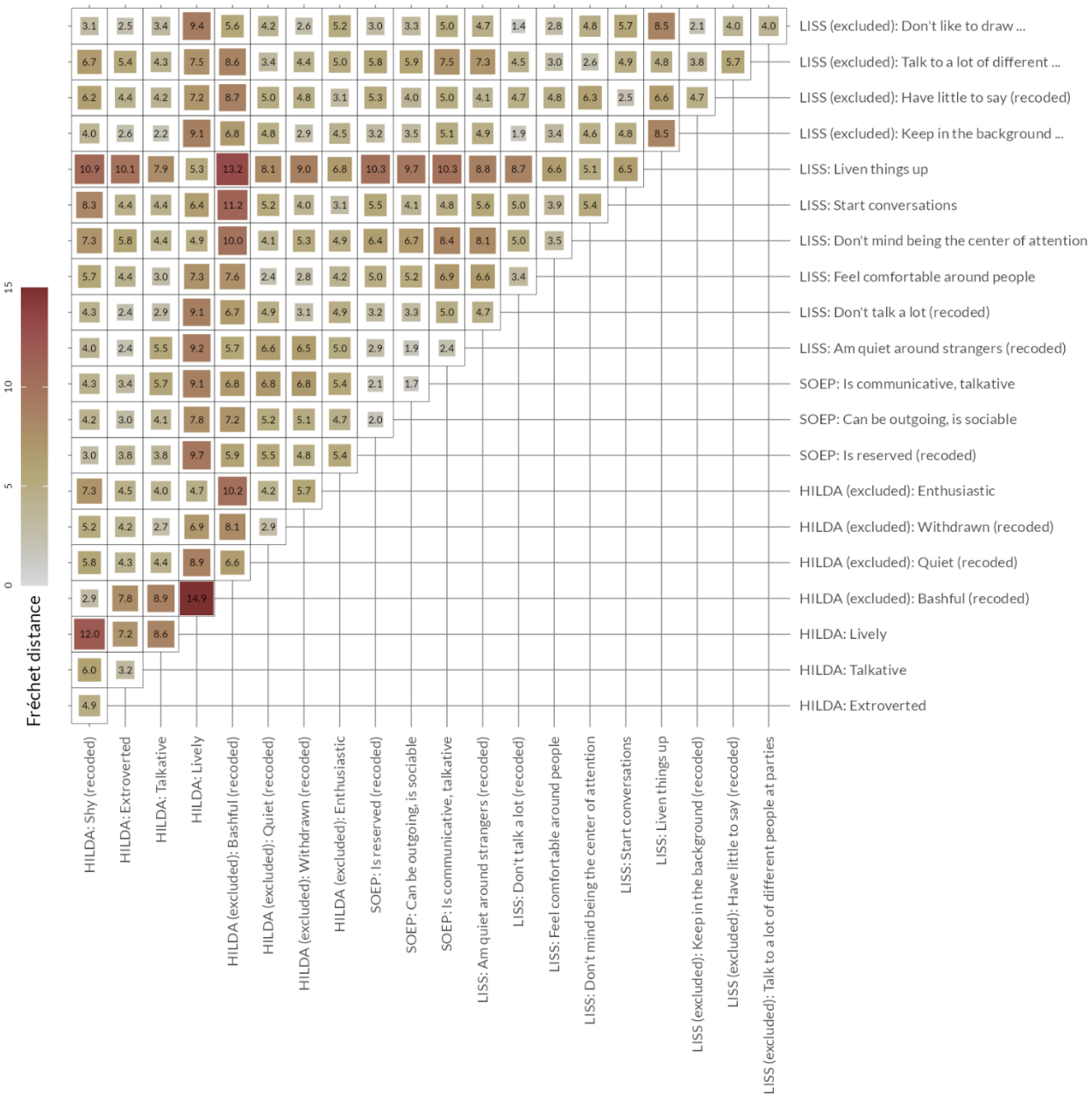
Similarity of the Within-Person Item-Level Age Trajectories for Neuroticism in the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the Socio-Economic Panel (SOEP), and the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. Items marked as "excluded" did not contribute to the trait mean score.

Figure S21

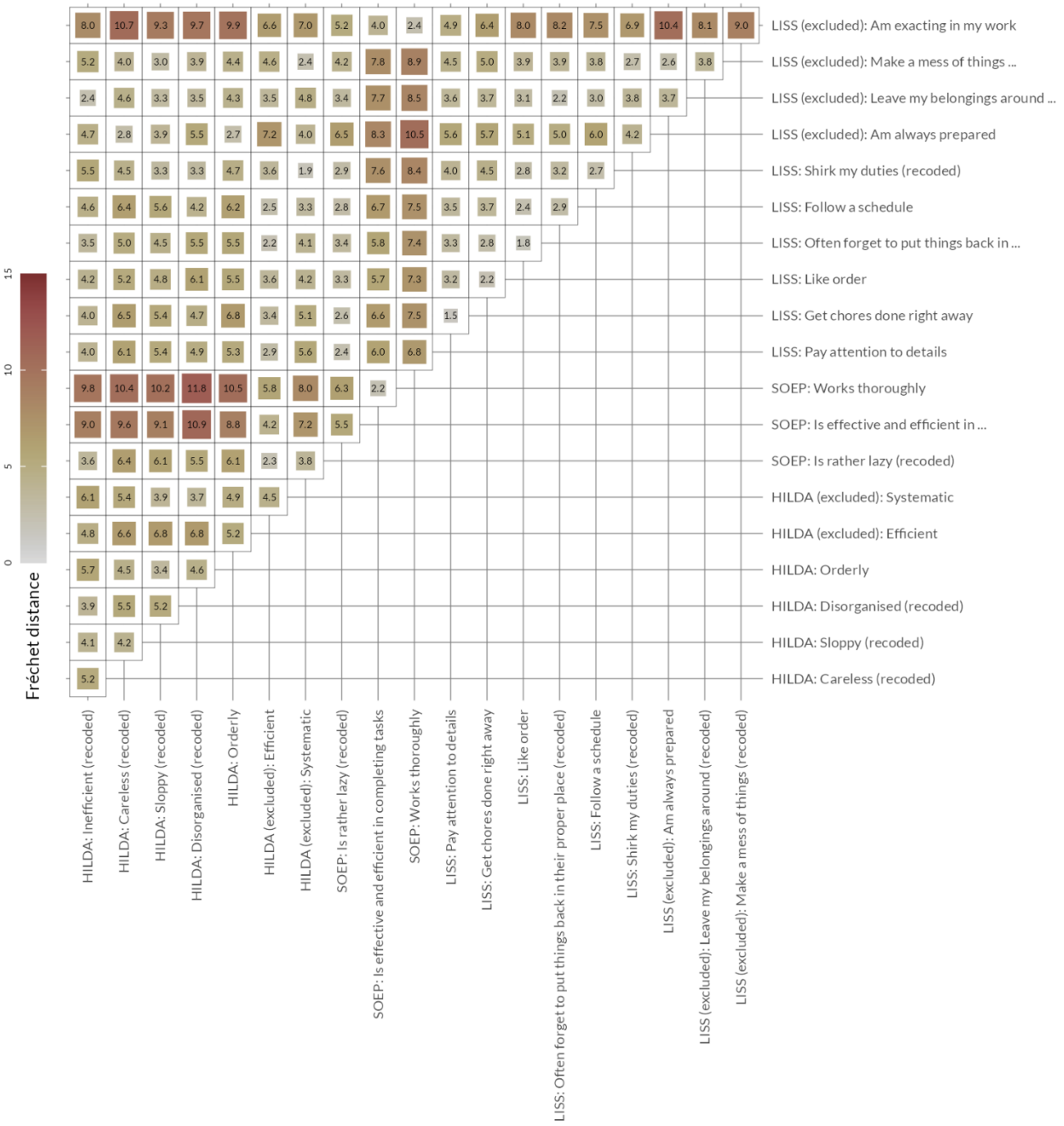
Similarity of the Within-Person Item-Level Age Trajectories for Extraversion in the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the Socio-Economic Panel (SOEP), and the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. Items marked as "excluded" did not contribute to the trait mean score.

Figure S22

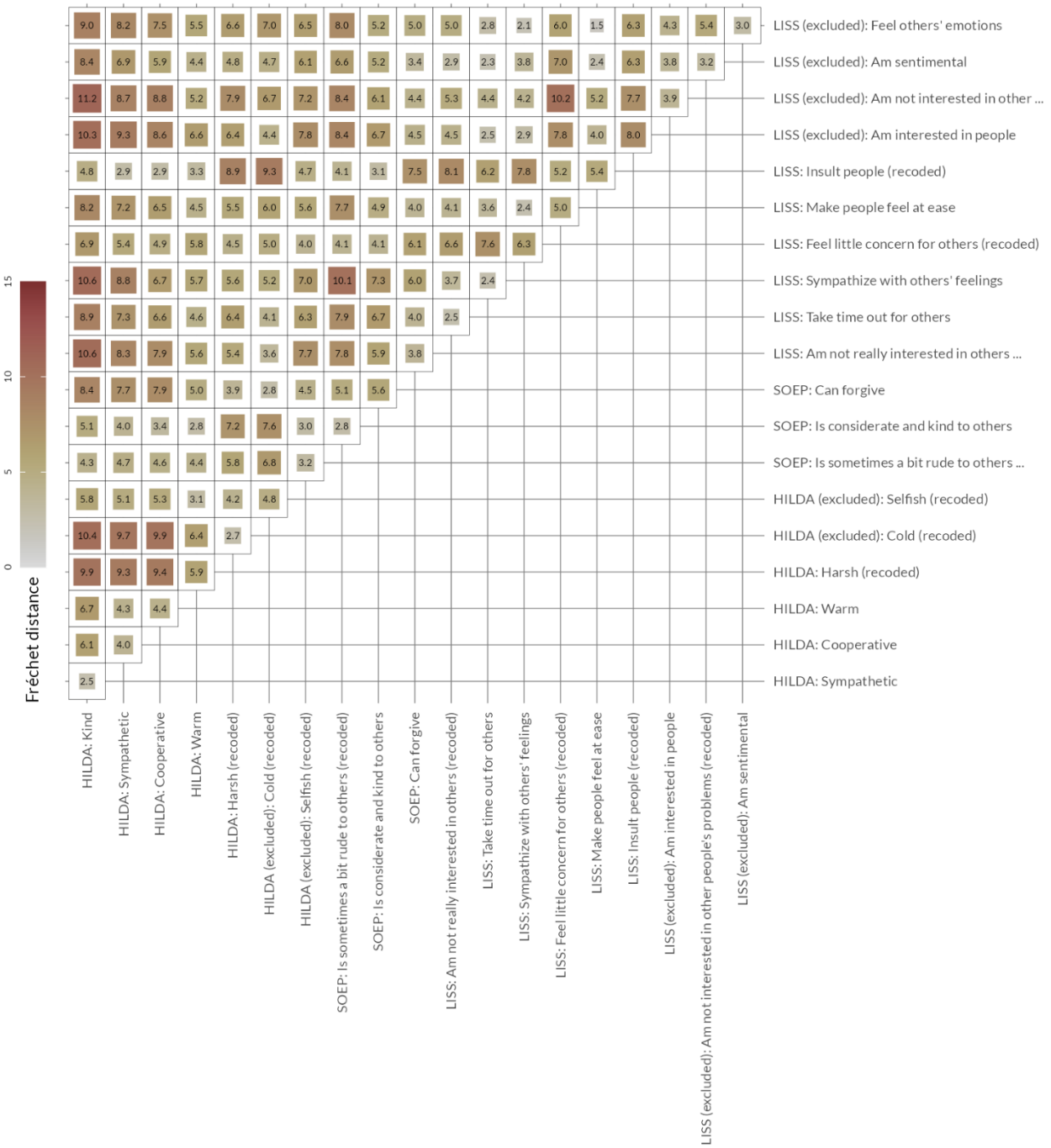
Similarity of the Within-Person Item-Level Age Trajectories for Conscientiousness in the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the Socio-Economic Panel (SOEP), and the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. Items marked as "excluded" did not contribute to the trait mean score.

Figure S23

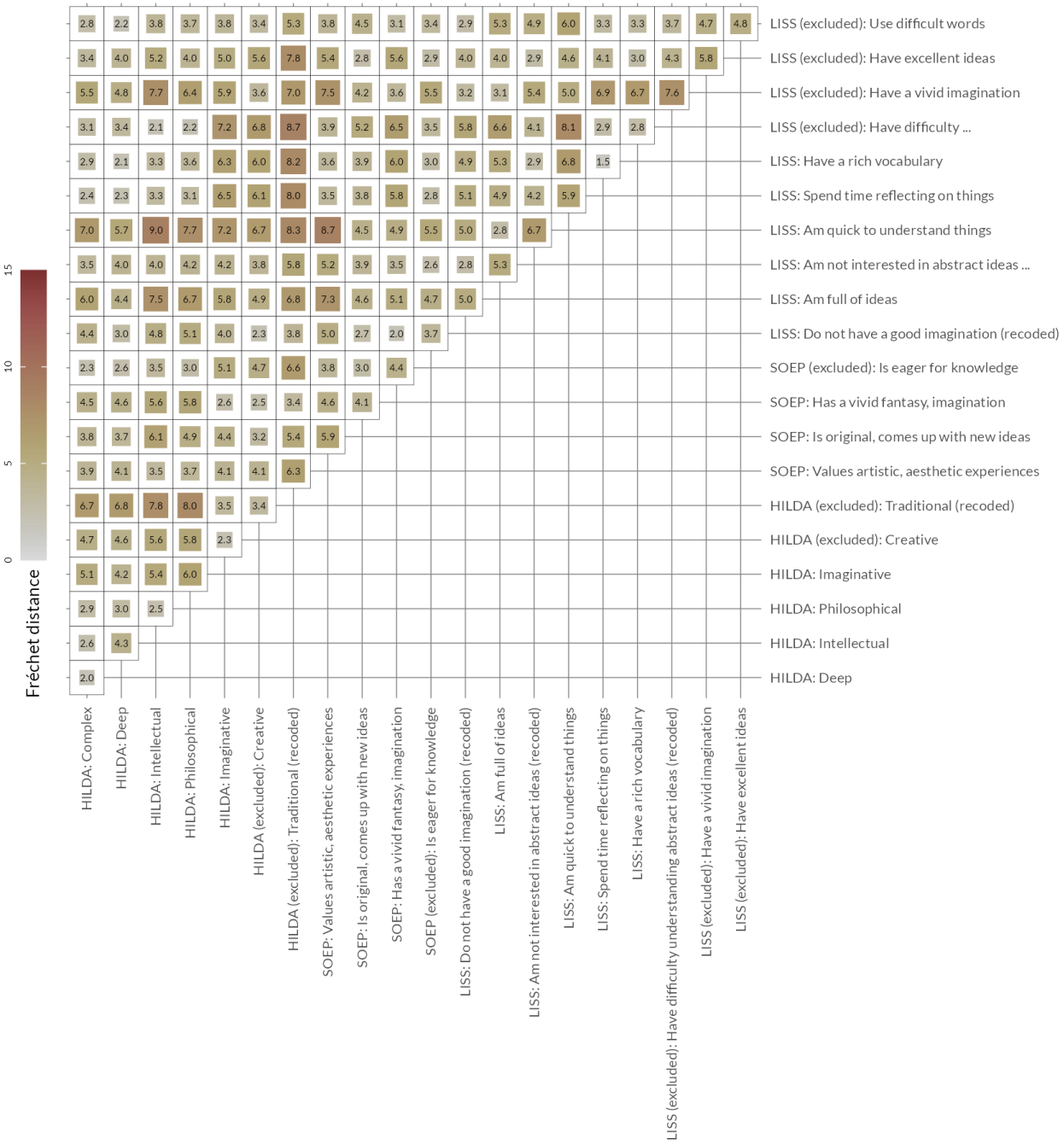
Similarity of the Within-Person Item-Level Age Trajectories for Agreeableness in the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the Socio-Economic Panel (SOEP), and the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. Items marked as "excluded" did not contribute to the trait mean score.

Figure S24

Similarity of the Within-Person Item-Level Age Trajectories for Openness in the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the Socio-Economic Panel (SOEP), and the Longitudinal Internet Studies for the Social Sciences (LISS) Panel



Note. Items marked as "excluded" did not contribute to the trait mean score.

Within-Person Factor Structure

For the present investigation, we relied on only within-person changes to estimate the age trajectories of the Big Five personality traits. However, the Big Five were derived in such a way that they primarily describe differences *between* individuals (e.g., Beck & Jackson, 2020; John, 2021). Thus, it is not fully clear whether the Big Five reflect a suitable framework for representing changes that occur *within* individuals. To investigate the matter, we tested whether the same latent model of personality held on both the between-person and within-person levels by means of multilevel confirmatory factor analysis (MLCFA; Geldhof et al., 2014; Muthén, 1994; see also Grosz, 2020).

In MLCFA, a latent factor model is estimated simultaneously for the between-person and within-person levels. For applications of MLCFA to longitudinal data, the between-person level represents time-invariant differences in the items between individuals. To obtain the item covariance matrix for the between-person level, an individual's item responses are averaged across the measurements (for estimation, this matrix is corrected for within-person differences; Muthén, 1994). By contrast, the within-person level represents average changes in the individuals' item responses across time. For the within-person level, the item covariance matrix is calculated by pooling the person-mean-centered item covariance matrices across individuals (Muthén, 1994).

For our analyses with MLCFA, we relied on the same data that we used to estimate the within-person age trajectories of personality scores in Studies 1 to 3. Separately for each trait and study, in MLCFA, we assumed the same unidimensional measurement model on both the between-person and within-person levels (i.e., configural invariance). For both levels, the same items constituted the manifest indicators of one latent factor (i.e., the corresponding personality

construct). No residual correlations were modeled, and no additional constraints were added to the measurement model.

We used maximum likelihood (ML) estimation for MLCFA. The overall model fit was evaluated with the CFI and RMSEA; the SRMR was assessed separately on the between-person and within-person levels. Despite there being some uncertainty regarding cutoffs for these fit indices in MLCFA (Hsu et al., 2015), we used conventional criteria (Hu & Bentler, 1999) to evaluate model fit (i.e., for an acceptable model fit, CFI $\geq$ .95, RMSEA $\leq$ .06, and SRMR $\leq$ .08).

In Study 2 (SOEP), because three items were used to measure each Big Five trait, model fit would be perfect without additional constraints (due to having just-identified models). To evaluate model fit for MLCFA in Study 2, we defined the factor loadings of each item to be the same on the between-person and within-person levels (i.e., metric invariance; resulting in overidentified models). Accordingly, the resulting fit indices represented a lower bound for model fit. Further, for one Agreeableness item in Study 2 (i.e., "Is considerate and kind to others"), we had to fix the residual variance to 0 in the between-person part to avoid having a negative model-implied item variance. Adding this constraint had a negligible impact on model fit.

The results from the MLCFA are reported in Table S19. With a minor exception (i.e., CFI for Conscientiousness in Study 3), the fit indices indicated a good model fit for each personality trait in the three studies. Thus, the analyses suggested that the same measurement model of a Big Five trait held on the between-person and within-person levels.

Table S19

Fit Indices for Testing for an Equal Factor Structure for the Big Five Across the Within-Person and Between-Person Levels

Model	*n*	Observations	χ^2 [a]	*df*	CFI	RMSEA	SRMR	
							Within	Between
Study 1: Household, Income and Labour Dynamics in Australia (HILDA) Survey								
Neuroticism	15,083	45,424	810.874	18	.983	.031	.013	.031
Extraversion	15,014	45,084	216.671	4	.988	.034	.016	.018
Conscientiousness	15,075	45,345	1,326.745	10	.969	.054	.020	.048
Agreeableness	15,168	45,744	274.385	10	.994	.024	.008	.021
Openness	15,009	45,132	980.981	10	.963	.046	.022	.039
Study 2: Socio-Economic Panel (SOEP) [b]								
Neuroticism	22,727	74,168	170.733	2	.991	.034	.019	.019
Extraversion	22,722	74,120	377.970	2	.986	.050	.031	.037
Conscientiousness	22,639	73,843	365.819	2	.985	.050	.025	.047
Agreeableness [c]	22,716	74,100	387.831	3	.970	.042	.030	.028
Openness	22,593	73,585	31.232	2	.998	.014	.008	.008
Study 3: Longitudinal Internet Studies for the Social Sciences (LISS) Panel								
Neuroticism	10,163	45,692	1,602.622	18	.962	.044	.020	.042
Extraversion	10,163	45,692	1,044.761	18	.971	.035	.017	.032
Conscientiousness	10,163	45,692	1,374.097	18	.935	.041	.021	.047
Agreeableness	10,163	45,692	1,283.160	18	.962	.039	.024	.030
Openness	10,163	45,692	819.490	18	.949	.031	.015	.042

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

[a] All *p*s < .001. [b] To allow for model evaluation, the factor loadings of each item were set equal across the within-person and between-person levels. [c] The residual variance of one item was fixed to 0 to avoid having a negative model-implied variance for the observed variables.

However, as the preceding analyses examined each personality trait in isolation, we additionally tested for whether the same factor structure held on both the between-person and within-person levels when the Big Five were all considered together. So, for each study separately, we integrated the five unidimensional measurement models from the previous analyses into a broader model with five latent factors (i.e., the Big Five traits). The latent factors were allowed to covary with each other. No residual correlations or cross-loadings were modeled; no additional constraints were added to the measurement models. For the analyses of the five-factor models in MLCFA, we included participants if they provided complete answers on all Big Five items in at least two waves of measurement in the respective study. Again, for one Agreeableness item in Study 2 (i.e., “Is considerate and kind to others”), we had to fix the residual variance to 0 in the between-person part to avoid having a negative model-implied item variance. Adding this constraint had a negligible impact on model fit. When there is a lack of good overall model fit in MLCFA, it might be difficult to indicate whether the source of misspecification is located on the between-person or within-person level, as the two parts are combined in one model. Therefore, we additionally examined the Big Five factor structure for the within-person and between-person levels in isolation by analyzing the corresponding covariance matrices in separate CFA models.

As Table S20 shows, assuming the Big Five factor structure simultaneously on the between-person and within-person levels resulted in a rather unsatisfactory general model fit for each study in MLCFA (especially indicated by CFI). However, the general misfit might not necessarily be due to issues exclusively on the within-person level but might already exist at the between-person level. Indeed, when we examined the Big Five factor structure in separate CFA models for the within-person and between-person levels, the model fit was unacceptable for the

between-person level in each study; but for the within-person level, the RMSEA and SRMR consistently indicated good fit, whereas the CFI indicated a lack of model fit (see Table S20). Thus, although it was not clear whether or not the Big Five factor structure fit the data on the within-person level, the discrepancy was definitely smaller than on the between-person level.

Table S20

Fit Indices for Testing for the Big Five Factor Structure

Model	*n*	Observations	χ^2 [a]	*df*	CFI	RMSEA	SRMR	
							Within	Between
Study 1: Household, Income and Labour Dynamics in Australia (HILDA) Survey								
MLCFA	14,554	43,330	60,298.254	530	.758	.051	.073	.138
Within-person model			24,616.172	265	.704	.057	.069	
Between-person model			64,514.925	265	.754	.129	.134	
Study 2: Socio-Economic Panel (SOEP)								
MLCFA [b]	22,293	72,103	30,633.201	161	.809	.051	.055	.099
Within-person model			13,899.065	80	.760	.059	.054	
Between-person model [b]			30,772.836	81	.816	.130	.101	
Study 3: Longitudinal Internet Studies for the Social Sciences (LISS) Panel								
MLCFA	10,163	45,692	40,934.881	790	.800	.033	.029	.110
Within-person model			11,341.119	395	.821	.028	.029	
Between-person model			51,988.793	395	.752	.113	.110	

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. MLCFA = Multilevel confirmatory factor analysis.

[a] All *p*s < .001. [b] The residual variance of one Agreeableness item was fixed to 0 to avoid having a negative model-implied variance for the observed variables.

Technical Appendix: Equivalence of Fixed Effects Modeling and Multilevel Modeling With Person-Mean Centering

To obtain pure within-person estimates from longitudinal data, FEM represents an easy-to-implement analytical approach. However, in psychological research, longitudinal data are typically analyzed with MLM, which in its standard implementation blends between-person and within-person effects in the estimates (Curran & Bauer, 2011). But, person-mean centering in MLM allows researchers to obtain within-person estimates that are identical to those from analyses with FEM (see, e.g., Hamaker & Muthén, 2020). To illustrate the equivalence of FEM and MLM with person-mean centering, we used the Neuroticism mean-score data from Study 2. The respective analysis script is publicly available on the OSF and can be accessed at https://osf.io/uycx6/.

To estimate the within-person age trend for Neuroticism in Study 2, we used dummy-coded age (ranging from 16 to 93, with age 16 as the reference group) as well as dummies for the first wave and mode, resulting in a total of 79 predictor variables. For FEM, the data could be entered directly into the model estimation process without any additional preparation. To obtain identical estimates in MLM, however, each of the 79 predictor variables first had to be individually person-mean centered. Similar to FEM, no intercept (representing the reference group) was included in MLM, and random intercepts (but no random slopes) were estimated for individuals. We used a restricted maximum likelihood (REML) estimator for MLM.

As Table S21 shows, the estimated slopes (and corresponding nonrobust standard errors) were identical between FEM and MLM when person-mean centering was applied. However, *without* person-mean centering, MLM provided quite different results (for a contrast of the resulting age trends, see Figure S17, as well as Figure 4 in the manuscript). Practically speaking,

FEM represents a straightforward way to obtain pure within-person effects, whereas MLM with person-mean centering might be more complicated to implement and error-prone just to yield the same results.

Table S21

Slope Estimates for the Neuroticism Mean Score in the Socio-Economic Panel (SOEP) Based on Fixed Effects Modeling (FEM) and Multilevel Modeling (MLM) With and Without Person-Mean Centering

Parameter	FEM		MLM			
			Person-mean centering		No person-mean centering	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Age dummy						
17	−1.132	0.605	−1.132	0.605	−0.716	0.514
18	−0.868	0.529	−0.868	0.529	−0.318	0.490
19	−1.138	0.596	−1.138	0.596	−0.565	0.503
20	−0.645	0.446	−0.645	0.446	0.104	0.427
21	−0.563	0.579	−0.563	0.579	0.562	0.488
22	−1.592	0.522	−1.592	0.522	−0.420	0.472
23	−1.047	0.578	−1.047	0.578	0.045	0.484
24	−0.931	0.503	−0.931	0.503	0.421	0.457
25	−0.818	0.578	−0.818	0.578	0.616	0.488
26	−1.680	0.541	−1.680	0.541	−0.268	0.476
27	−1.543	0.575	−1.543	0.575	−0.046	0.480
28	−2.182	0.534	−2.182	0.534	−0.341	0.467
29	−1.514	0.579	−1.514	0.579	0.379	0.481
30	−2.001	0.557	−2.001	0.557	−0.037	0.475
31	−2.069	0.579	−2.069	0.579	−0.089	0.476
32	−2.481	0.564	−2.481	0.564	−0.120	0.473
33	−2.193	0.585	−2.193	0.585	0.055	0.473
34	−2.703	0.576	−2.703	0.576	−0.166	0.473
35	−2.338	0.589	−2.338	0.589	0.092	0.469
36	−2.953	0.582	−2.953	0.582	−0.221	0.467
37	−2.909	0.596	−2.909	0.596	−0.205	0.467
38	−2.950	0.587	−2.950	0.587	0.043	0.461
39	−2.958	0.603	−2.958	0.603	−0.043	0.464
40	−3.311	0.595	−3.311	0.595	−0.050	0.458
41	−3.252	0.611	−3.252	0.611	−0.088	0.463
42	−3.670	0.606	−3.670	0.606	−0.206	0.458
43	−3.929	0.616	−3.929	0.616	−0.484	0.459
44	−4.141	0.612	−4.141	0.612	−0.306	0.453
45	−3.998	0.621	−3.998	0.621	−0.342	0.455
46	−4.147	0.620	−4.147	0.620	−0.181	0.454
47	−4.071	0.627	−4.071	0.627	−0.121	0.454
48	−4.547	0.628	−4.547	0.628	−0.244	0.452
49	−4.021	0.635	−4.021	0.635	0.057	0.452
50	−4.589	0.636	−4.589	0.636	−0.036	0.452
51	−4.636	0.643	−4.636	0.643	−0.131	0.453
52	−5.196	0.645	−5.196	0.645	−0.320	0.454
53	−4.707	0.650	−4.707	0.650	−0.046	0.451
54	−4.777	0.655	−4.777	0.655	0.318	0.454
55	−5.049	0.659	−5.049	0.659	−0.012	0.452
56	−5.360	0.664	−5.360	0.664	0.194	0.455
57	−5.266	0.669	−5.266	0.669	0.085	0.454
58	−5.573	0.673	−5.573	0.673	0.031	0.456

Parameter	FEM		MLM			
			Person-mean centering		No person-mean centering	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
59	−5.462	0.682	−5.462	0.682	0.183	0.458
60	−6.475	0.685	−6.475	0.685	−0.313	0.458
61	−6.044	0.689	−6.044	0.689	−0.032	0.456
62	−6.607	0.694	−6.607	0.694	−0.216	0.460
63	−6.708	0.698	−6.708	0.698	−0.374	0.457
64	−6.995	0.705	−6.995	0.705	−0.197	0.459
65	−6.835	0.709	−6.835	0.709	−0.171	0.457
66	−7.958	0.718	−7.958	0.718	−0.802	0.462
67	−7.614	0.722	−7.614	0.722	−0.558	0.461
68	−7.780	0.729	−7.780	0.729	−0.396	0.464
69	−7.677	0.733	−7.677	0.733	−0.375	0.462
70	−8.102	0.741	−8.102	0.741	−0.498	0.467
71	−7.921	0.751	−7.921	0.751	−0.263	0.473
72	−7.794	0.755	−7.794	0.755	−0.022	0.474
73	−7.375	0.766	−7.375	0.766	0.455	0.480
74	−7.976	0.770	−7.976	0.770	−0.009	0.480
75	−7.999	0.774	−7.999	0.774	0.067	0.483
76	−8.082	0.784	−8.082	0.784	0.046	0.490
77	−8.154	0.792	−8.154	0.792	0.069	0.492
78	−8.921	0.805	−8.921	0.805	−0.453	0.503
79	−8.582	0.815	−8.582	0.815	−0.012	0.512
80	−8.355	0.832	−8.355	0.832	0.204	0.528
81	−9.014	0.844	−9.014	0.844	−0.339	0.538
82	−9.044	0.862	−9.044	0.862	−0.302	0.555
83	−8.878	0.892	−8.878	0.892	0.128	0.586
84	−8.991	0.921	−8.991	0.921	−0.256	0.615
85	−9.019	0.951	−9.019	0.951	0.247	0.648
86	−10.131	1.004	−10.131	1.004	−1.144	0.712
87	−8.422	1.016	−8.422	1.016	0.638	0.709
88	−9.734	1.092	−9.734	1.092	−0.662	0.792
89	−9.953	1.161	−9.953	1.161	−0.532	0.880
90	−9.313	1.307	−9.313	1.307	0.393	1.051
91	−7.593	1.343	−7.593	1.343	1.586	1.054
92	−7.916	1.682	−7.916	1.682	1.169	1.418
93	−10.062	1.650	−10.062	1.650	−0.521	1.390
First wave	0.261	0.084	0.261	0.084	1.165	0.059
Mode	1.140	0.114	1.140	0.114	1.261	0.086

Note. The first wave was dummy-coded as 1 = first wave and 0 = all other waves. Mode was dummy-coded as 1 = self-report questionnaire and 0 = interview.

References

Alt, H., & Godau, M. (1995). Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, *5*(1–2), 75–91. https://doi.org/10.1142/S0218195995000064

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434. https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x

Beck, E. D., & Jackson, J. J. (2020). Idiographic traits: A return to Allportian approaches to personality. *Current Directions in Psychological Science*, *29*(3), 301–308. https://doi.org/10.1177/0963721420915860

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Fréchet, M. M. (1906). Sur quelques points du calcul fonctionnel [On some points of the functional calculus]. *Rendiconti Del Circolo Matematico Di Palermo*, *22*(1), 1–72. https://doi.org/10.1007/BF03018603

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Genolini, C., Ecochard, R., Benghezal, M., Driss, T., Andrieu, S., & Subtil, F. (2016). kmlShape: An efficient method to cluster longitudinal data (time-series) according to their shapes. *PLoS ONE*, *11*(6), Article e0150738. https://doi.org/10.1371/journal.pone.0150738

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Grosz, M. P. (2020). *The factor structure of Big Five personality trait measures at the between- and within-person levels*. PsyArXiv. https://doi.org/10.31234/osf.io/k6r7g

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Hsu, H.-Y., Kwok, O.-M., Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research*, *50*(2), 197–215. https://doi.org/10.1080/00273171.2014.977429

Hu, L., & 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. https://doi.org/10.1080/10705519909540118

John, O. P. (2021). History, measurement, and conceptual elaboration of the Big-Five trait taxonomy: The paradigm matures. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (4th ed., pp. 35–82). Guilford Press.

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

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful tools for structural equation modeling* (Version 0.5-5) [R package]. https://CRAN.R-project.org/package=semTools

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*(3), 376–398. https://doi.org/10.1177/0049124194022003006

Olaru, G., Schroeders, U., Wilhelm, O., & Ostendorf, F. (2019). 'Grandpa, do you like roller coasters?': Identifying age-appropriate personality indicators. *European Journal of Personality*, *33*(3), 264–278. https://doi.org/10.1002/per.2185

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Seifert, I. S., Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2022). The development of the rank-order stability of the Big Five across the life span. *Journal of Personality and Social Psychology*, *122*(5), 920–941. https://doi.org/10.1037/pspp0000398

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences*, *115*(1), E15–E23. https://doi.org/10.1073/pnas.1712277115