European Journal of Personality
Eur. J. Pers. **15**: 169–198 (2001)
DOI: 10.1002/per.408

Carving Personality Description at its Joints: Confirmation of Three Replicable Personality Prototypes for Both Children and Adults

JENS B. ASENDORPF[1]*, PETER BORKENAU[2], FRITZ OSTENDORF[3] and MARCEL A.G. VAN AKEN[4]

[1]*Humboldt-Universität zu Berlin, Germany*
[2]*Martin-Luther-Universität Halle-Wittenberg, Germany*
[3]*Universität Bielefeld, Germany*
[4]*University of Nijmegen, The Netherlands*

Abstract

We tested the hypothesis that there are three major prototypic patterns of personality description (resilient, overcontrolled, and undercontrolled) in a series of studies including adults' self-descriptions on the Big Five and parents' Big Five and Q-Sort judgments of their children, using both replicated cluster analyses and replicated Q-factor analyses. The consistency of the prototypes across ages, judges, and methods was quantitatively measured. The results confirmed the hypothesis in all studies. Personality, social relationship, and social interaction correlates of the prototypes indicated externalizing tendencies for undercontrollers and internalizing tendencies for overcontrollers for both children and adults. The studies provide strong evidence for a three-prototype model of personality description at the highest level of analysis for both childhood and adulthood.

INTRODUCTION

Personality can be defined as 'the dynamic organization *within the individual* of those psychophysical systems that determine his unique adjustments to his environment' (Allport, 1937, p.48). Thus, personality psychology is concerned with the description, prediction, and explanation of this within-person organization, or personality structure. The unit of analysis for these tasks is the individual person. The structure of personality can be studied within a trait perspective on personality (Allport, 1937; Funder, 1991). Psychologically meaningful characteristics on which individuals reliably differ (traits) are isolated, and an individual's personality is described by a pattern (configuration) of traits. This description should be both broad (covering a wide variety of different traits) and

*Correspondence to: Jens B. Asendorpf, Institut für Psychologie, Humboldt-Universität zu Berlin, Oranienburger Strasse 18, 10178 Berlin, Germany. E-mail: asen@rz.hu-berlin.de

Received 27 July 2000
Accepted 30 January 2001

efficient (representing that wide variety of traits by a minimal set of trait dimensions). The five-factor model of personality description provides one solution to this problem at a high level of analysis (few trait dimensions that cover a wide variety of traits). The five-factor model can be applied to the description of adults', adolescents', and children's personality patterns (Digman, 1990; John, 1990; Halverson, Kohnstamm, & Martin, 1994). It can be differentiated into a more fine-grained system of personality description in terms of facets (subfactors) of these 'Big Five' factors (see McCrae and Costa, 1992; Saucier and Ostendorf, 1999).

CARVING PERSONALITY DESCRIPTION AT ITS JOINTS

Whereas there has been, and still is, considerable debate about the question whether the 'Big Five' are an optimal set of trait dimensions for descriptive and predictive purposes (see particularly Block, 1995, and Paunonen, 1998), another question has been much less discussed and studied empirically. Deriving an efficient set of trait dimensions, and representing personality structure as a point in the multivariate trait space, is not sufficient for the description of personality. It is not sufficient because we do not know how the individual personality patterns are distributed in the multivariate trait space. If the trait dimensions do not correlate as in the case of orthogonal trait factors, this linear independence does not imply that the dimensions are statistically independent. Stated differently, the dimensions may show substantial nonlinear interactions, not only in terms of quadratic or higher-order curvilinear relationships but in terms of any uneven filling of the multivariate trait space. As Gangestad and Snyder (1985) put it, we can 'carve nature at its joints' if we treat uneven multivariate distributions with respect. Because we are dealing here with verbal descriptions of personality rather than behavioral assessments, we should more cautiously say that we can 'carve personality description at its joints'.

Because it is difficult to visualize multivariate distributions for more than two dimensions, we discuss the problem of uneven multivariate distributions for the most simple case of only two trait dimensions *X* and *Y*. Thus, personality patterns in this case are described by pairs of scores (x,y). However, our discussion can be fully generalized to any trait space with more than two dimensions.

In Figure 1, four different kinds of bivariate distribution of *X* and *Y* are presented; in all four cases, the correlation between *X* and *Y* is zero. Figure 1(A) presents the ideal case of a multivariate linear model, the binormal distribution. The patterns centre around the means of *X* and *Y*. The more distant the trait scores x and y of an individual are from this centre of the distribution, the less frequent they are. The univariate distributions of the variables *X* and *Y*, but also of any rotation of them in the bivariate space, are Gaussian bell-shaped curves. In this case, *X* and *Y* are statistically independent, i.e. nothing can be gained for the prediction of y from knowing x, and vice versa. It is possible, of course, to classify the personality patterns in this case into classes such as high *X* or high *X*–low *Y* according to cut-points along *X* and *Y*, but these classes are completely arbitrary. Whereas factor scores of multidimensional personality questionnaires often come close to the ideal of a multinormal distribution, the raw scale scores often deviate from such a distribution.

Consider now Figure 1(B). Again, *X* and *Y* do not correlate, but they show a very different bivariate distribution. Neither a linear model adequately represents the observed personality differences, nor a distinction between highs and lows in *X*, *Y*, or *X* and *Y*. There

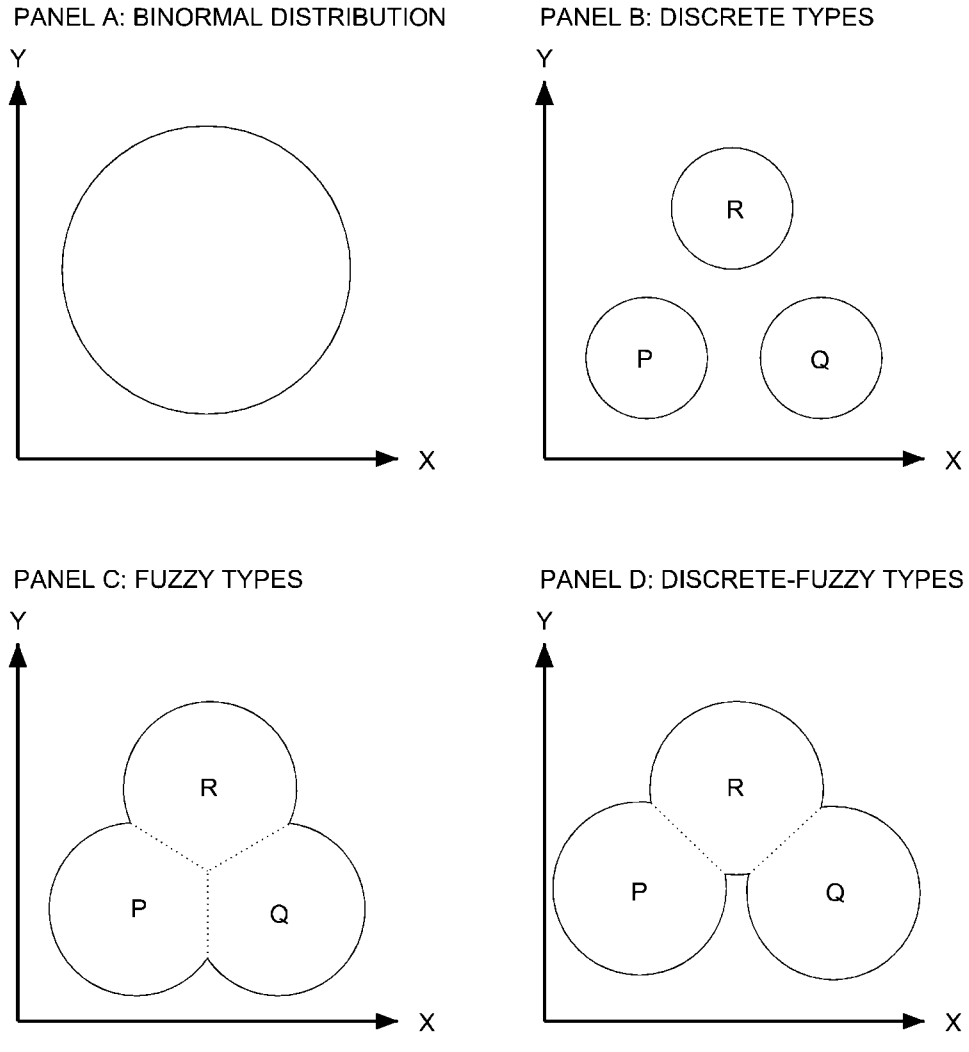

Figure 1. Schematic presentation of four bivariate distributions of uncorrelated trait dimensions.

is a quadratic relationship between *X* and *Y* because high scores on *Y* are incompatible with extreme scores on *X*, but this curvilinear model also does not adequately represent the data. Instead, people cluster around three prototypic personality patterns P, Q, and R. If we define a personality type as a group of individuals whose personality patterns are similar to such a personality prototype, the whole population can be sorted into three clearly distinguishable types. In such a prototype model, members of each type do not show the same personality pattern, but vary gradually with regard to their prototypicality (their similarity with the best-fitting prototype, graphically represented in Figure 1(B) by the distance from the prototype). However, every person can be unambiguously assigned to one of the three types because the borders of the types are clearly separated. In other words, the transition from one type to the next is discontinuous; the types are discrete.

It would be surprising if personality patterns were distributed according to this ideal typological model because this would imply that there exist a few highly potent genetic or

environmental causes of personality (or highly correlated bundles of such causes) that group all members of a population into clearly separable types. Stereotyping of personality differences in personality judgment may increase the likelihood of finding prototypic organizations of personality patterns. However, inspection of the bivariate distributions of pairs of Big Five scales does not show evidence for discrete types.

Unfortunately, discussions of a continuous-dimensional versus a categorical–typological approach to personality nearly always contrast the two unrealistic cases (A) and (B) with one another. More realistic is the assumption that the borders between the types are fuzzy because the types are not clearly separated. In this case, nearly all people can still be assigned to a best-fitting prototype, but small variations in a personality pattern, which may be only due to unreliable measurement, change the type membership. In other words, there is an area of uncertainty of classification at the borders between the types. These areas are graphically presented in Figure 1(C) as fused circles around the prototypes. The transition between the types is smooth; they are fuzzy types. Fuzziness in this sense does not imply impreciseness of description or even inaccessibility to science (see e.g. Zetenyi, 1988, for the mathematical treatment of fuzzy sets and its application to psychology).

When Figures 1(A)–(C) are contrasted with one another, it becomes obvious that there is a gradual transition between the multivariate distribution (A) where all types are completely fused, via the fuzzy type model (C) to the discrete type model (B) where all types are completely discrete. Thus, (A) and (B) can be considered extreme cases of the more general fuzzy type model.

Recently, Asendorpf and van Aken (1999) showed that personality patterns can be distributed in another way that is partly discrete and partly fuzzy. In their analysis of children's personality patterns they found a distribution as illustrated by Figure 1(D). From a dimensional point of view, the continuous dimension *Y* bifurcates in its lower part into two groups of low-*Y* scorers that are discrete relative to one another with regard to dimension *X*, but continuous with high *Y*. From a typological point of view, three different personality prototypes were identified. The P and R types, and the Q and R types, were fused whereas the P and Q types were discrete relative to another. In other words, the transition from P to Q was discontinuous, but the transitions from R to P and Q were smooth. The discontinuity of the transition between P and Q was empirically identified by bimodal distributions of the prototypicalities for P and Q when the R type was dropped from analysis. This case violates both a purely continuous and a purely categorical view of the distribution of personality patterns. However, there is no problem in describing it as one special case of the fuzzy type model.

Thus, one approach of studying uneven multivariate distributions of personality patterns is a type approach that acknowledges (a) the prototypical organization within types, and (b) fuzzy borders between types. If fuzzy types can be reliably distinguished, such a reconstruction of the distribution of personality patterns has four main advantages. First, it offers an empirically based, more adequate description of personality differences than the traditional dimensional trait model that explicitly assumes linear independence, but implicitly also statistical independence. Second, the fact that particular traits tend to co-occur in a type tells us something about the within-individual organization of personality. It carves personality description at its joints. In contrast, a linear independence model is silent about this organization. Third, the detection of such 'joints' may provide clues about genetic or environmental causes of personality structure that are not noticed by correlations between trait dimensions and explanatory variables. And fourth, personality types, fuzzy or discrete, offer a nonarbitrary classification system for personality structure.

Classifying people by cut-points along one or two trait dimensions leads to arbitrary classifications. Types are nonarbitrary cuts of the full multivariate trait space.

HOW CAN PERSONALITY PROTOTYPES BE EMPIRICALLY IDENTIFIED?

The multinormal distribution can be considered as the case of only one personality type that is characterized by the mean personality pattern in the population (see Figure. 1). Thus, the crucial question is whether the population can be reliably split into multiple types, and into how many. Any answer to this question is likely to depend on (a) the trait dimensions on which the patterns are based and the operationalization of the traits (e.g. by self-rating or behavioural observation), (b) the population (e.g. culture, age), and (c) the method of deriving empirical types from the distribution of the patterns in the population. For example, other things being equal, more types may be found when the patterns are based on many instead of a few traits, and different types may be found for different sets of trait dimensions.

Particularly important is the problem of how chance findings are avoided that capitalize on noise in the data. For example, cluster analysis always generates clusters at any level of differentiation regardless whether the cluster differences are reliable or not. Our solution of this problem is to require that the types are both replicable within studies across random splits of the sample of participants, and generalizable across different ages and methods. Replicable and generalizable types indicate unevenness of the multivariate distribution; they cannot be found for multinormal distributions.These requirements prevent chance findings as well as findings that are specific to particular samples, ages, and methods.

Furthermore, the consistency across methods of deriving the types is important because two different methods exist. The first method was developed by J. Block (1961, 1971) and is based on a Q-factor analysis of Q-sort patterns. First, each individual is described by a knowledgeable informant who sorts trait descriptions according to how well they fit the individual's personality (Q-sort). Second, these individual Q-sort profiles are factor analysed by Q-factor analysis (also called inverse factor analysis because the roles of persons and variables in ordinary factor analysis are reversed). The resulting Q-factors represent prototypic personality patterns. Third, the individuals are classified into personality types according to the best-fitting Q-factor. As correlations between personality profiles (and the factors derived from them) ignore between-person differences in overall level and scatter (the intraindividual means and variances), this method should only be applied to Q-sort measures such as the California Adult Q-Sort (Block, 1971) and the California Child Q-Sort (Block and Block, 1980) that force the judges to produce Q-sorts with a prescribed mean and variance.

Applications of this method vary with regard to three main points: which set of Q-sort items, how many Q-factors, and how the Q-profiles are assigned to the Q-factors. Block (1971), York and John (1992), and Klohnen and Block (1995, unpublished analyses cited by Caspi, 1998) used the 100-item California Adult Q-Set and the 104-item California Adolescent Q-Set; Robins, John, Caspi, Moffitt and Stouthamer-Loeber (1996) and Hart, Hofmann, Edelstein and Keller (1997) used the 100-item California Child Q-Set. The items of these Q-sets were originally selected by J. Block and colleagues to cover a wide range of age-appropriate traits. John, Caspi, Robins, Moffitt and Stouthamer-Loeber (1994) showed that the Big Five factors of personality description are well represented by the items of the child Q-set; the same applies to the adult Q-set (McCrae, Costa, & Busch, 1986).

The number of Q-factors can be decided by inspection of the scree plot, which however offers only a relative criterion and often does not provide unequivocal information. A stricter criterion that can be also directly compared between studies is the replicability of the factors when the sample is split into two random halves (Everett, 1983). This criterion was first applied to Q-factor analysis by York and John (1992), and was used in all later Q-type studies. Using this criterion, both Robins *et al.* (1996) and Hart *et al.* (1997) found that three, but not more, Q-factors were replicable in their studies of children.

Most studies assigned individuals to Q-types on the basis of their highest Q-factor loading (in Q-factor analysis, there are factor loadings for persons and factor scores for variables). In some cases, problems arise because a Q-sort loads equally high on two factors, or low on all factors. Thus, the Q-types are fuzzy. To solve this problem, Robins *et al.* (1996) introduced a more complex procedure. First, only persons are classified that have a particularly high factor loading on one Q-factor. Second, the remaining persons are assigned to these Q-types by discriminant analysis.

Another approach to empirically deriving personality types is the cluster analysis of multiple trait scores (questionnaire scales, ratings, or test scores). Each individual is described by the profile of scores in these variables. These profiles are grouped by cluster analysis into relatively homogeneous clusters. Each cluster represents a personality type, and the average profile of the cluster members describes a personality prototype. Profiles that differ with regard to overall level and scatter can be treated as different if an appropriate measure of profile similarity is used (e.g., the Euclidean distance of the profiles). Again, applications of this method vary with regard to three main points: which trait assessments, which method of cluster analysis, and how many clusters.

Caspi and Silva (1995) obtained 22 behavioural ratings from examiners who observed the participating children in various testing situations, reduced these ratings by factor analysis to three factors, and clustered the resulting profiles of factor scores. van Lieshout, Haselager, Riksen-Walraven and van Aken (1995) clustered Big Five scales that were derived from the California Child Q-Set. In a study of female adults, Pulkkinen (1996) clustered 12 composites that comprised a total of 62 personality variables.

In all three studies, Ward's hierarchical clustering procedure was used. Because clusters that are fused in one step remain together in all later steps, this procedure can result in nonoptimal solutions. Therefore, Caspi and Silva (1995) optimized the final solution by an additional iterative procedure that sorts all participants once again to the best-fitting Ward cluster, recomputes the cluster centres, and repeats these steps until the cluster centres change only minimally. This so-called *k*-means procedure minimizes within-cluster differences.

Similarly to Q-factor analysis, the replicability of the clusters between random splits of the sample can be considered as a fairly strict criterion for the optimal number of clusters. Caspi and Silva (1995) used this criterion. However, they relied on a visual inspection of the similarity of the cluster centres, which does not seem to be wholly satisfactory. In the present study we used the straightforward, quantitative method of evaluating cluster replicability that was proposed by Breckenridge (1989).

CASPI'S CONJECTURE: THREE MAJOR TYPES

In an informal review of the results of the above five Q-factor studies and the three clustering studies, Caspi (1998) concluded that despite the different methods of deriving

types, the different ages, birth cohorts, raters, and sets of traits, and despite the fact that the number of types varied across the studies, three types showed a high similarity across studies. From this result, Caspi (1998) inferred the conjecture that these three types may constitute a core set of types for any generalizable personality typology.

Furthermore, Asendorpf and van Aken (1999) were able to replicate the findings of Robins *et al.* (1996) and Hart *et al.* (1997) of exactly three replicable Q-types for parental judgments of US boys and expert ratings of Icelandic children for still other judges (teachers), age groups within childhood, and languages (German and Dutch). Therefore the present research was guided not only by Caspi's (1998) conjecture that the three types would be found in all of our studies, but also by our stronger main hypothesis that these and only these three types would be replicable.

The three types were labelled as resilients, overcontrollers, and undercontrollers in the majority of the nine studies (the eight reviewed by Caspi, 1998, and the study by Asendorpf and van Aken, 1999). These labels refer to the theory of ego-control and ego-resiliency by J. H. Block and J. Block (1980). In this dimensional model of personality, ego-resiliency refers to the tendency to respond flexibly rather than rigidly to changing situational demands, particularly stressful situations. Ego-control refers to the tendency to contain versus express emotional and motivational impulses (strong control versus weak control).

J. H. Block and J. Block (1980) assumed that both extremely high and low ego-control are related to low ego-resiliency. Thus, three rather than 2×2 distinguishable types are expected. This expectation was supported by Asendorpf and van Aken (1999), who found for children an inverted U-shaped relation between their prototypicality for resiliency and their prototypicalities for over- and undercontrol. Asendorpf and van Aken (1999) also studied the distribution of the prototypicalities for the three prototypes and found a pattern conforming to Figure 1(D), which suggested a mixed discrete–fuzzy model for the types. Therefore, we explored in the present study whether this model can be generalized to other ages and sets of personality descriptors.

Robins *et al.* (1996) described the three childhood personality types in terms of the five-factor model of personality description. Following the procedure proposed by John *et al.* (1994), they defined Big Five scales on the basis of the items of the California Child Q-Set and studied the three personality types in terms of these five scales. In all five scales, resilients had slightly more socially desirable scores (e.g., above-average scores in emotional stability). Overcontrollers had particularly low scores in extraversion and emotional stability, and undercontrollers had particularly low scores in conscientiousness and agreeableness. This pattern was replicated by Asendorpf and van Aken (1999), who used an independent method of measuring the Big Five. For parsimony, we expected the same pattern also for the three personality types in adulthood.

CORRELATES OF THE THREE TYPES

External correlates of the three types were reported in all nine studies. The findings can be summarized by referring to the concepts of general psychological adjustment and internalizing versus externalizing problems introduced by Achenbach and Edelbrock (1981) for children and adolescents. Resilients were consistently found to be the largest group, and socially and cognitively well adjusted. Overcontrollers and undercontrollers showed tendencies for maladjustment, although this term should be applied with caution

because these two types represented relatively large sections of the population (approximately 25% in each case). Overcontrollers showed internalizing tendencies (e.g. more inhibited and shy, lower social self-esteem, more loneliness), and undercontrollers showed externalizing tendencies (e.g. higher rate of antisocial behavior, lower peer popularity). We prefer to use the term 'tendencies' instead of 'problems' because of the relatively large size of the overcontrolled and the undercontrolled types.

In a follow-up study of the children who were identified by Caspi and Silva (1995) at age 3 as resilient, overcontrolled, and undercontrolled, Caspi and Silva (1995) and Caspi, Moffitt, Newman and Silva (1996) found that overcontrollers and undercontrollers continued to show internal versus external tendencies into adulthood. Together, these results provide a consistent picture of resilients being adjusted, overcontrollers showing internalizing tendencies, and undercontrollers showing externalizing tendencies, independent of their age.

THE PRESENT RESEARCH

The present research is an attempt to confirm Caspi's conjecture of three generalizable types not only for children but also for adults, to confirm our main hypothesis that these and only these three types are replicable, and to study the correlates of the three types. We (a) compared three-, four- and five-prototype solutions for replicability in both childhood and adulthood, using both the Q-sort and the cluster approach, (b) evaluated the consistency of the replicable types across ages and methods by a quantitative coefficient of consistency, (c) evaluated the stability of the types over 18 months in adulthood to obtain an estimate of their retest reliability, and (d) studied social correlates of the types in both childhood and adulthood. In addition, we explored whether the mixed discrete–fuzzy model for Q-types in childhood of Asendorpf and van Aken (1999) can be generalized to other ages and methods of personality description. For these purposes, we reanalysed four German data sets: the normative NEO-FFI sample (Borkenau and Ostendorf, 1993), a study of bipolar adjective markers of the Big Five (Ostendorf, 1990), the Berlin Relationship Study (Asendorpf and Wilpers, 1998, 2000), and the LOGIC study (Weinert and Schneider, 1999).

METHOD

We organize this section by study, followed by a description of how we derived personality prototypes.

Adult NEO-FFI study

The participants of this study were part of the German NEO-FFI normative sample that comprised 2112 participants. The German NEO-FFI (Borkenau and Ostendorf, 1993) is a translation of the NEO Five-Factor-Inventory of Costa and McCrae (1989) that measures each of the factors Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness with 12 items. Of the 2112 participants, 195 had missing values and were therefore excluded. We further excluded all participants who were younger than 18 or older than 24 years to equalize the age of the participants in (a) the Adult NEO-FFI Study,

(b) the Adult Adjective Study, and (c) the Berlin Relationship Study. This left us with 365 male and 493 female participants with an average age of 21.6 years (SD = 1.58 years). Finally, to equalize the sex ratio in our sample, we excluded 128 randomly selected females, resulting in a final sample of 730 student-aged participants (365 females and 365 males) without any missing values. The internal consistencies of the five scales ranged from 0.72 for Openness to experience to 0.86 for Conscientiousness. The factor structure of the NEO-FFI turned out as expected: in a varimax-rotated principal component analysis of the 60 NEO-FFI items with the number of factors fixed to five, each item had its highest loading on the expected factor.

Adult adjective study

The participants of this study were part of a larger sample of 1928 participants aged 14–85 years who were recruited through advertisements in local newspapers, by psychology students who asked acquaintances for their participation, and in psychology courses. The participants provided self-ratings for 179 bipolar adjectives on bipolar six-point scales. Only participants with less than 10% missing values were included, and these values were estimated through regression. The 179 adjective pairs were marker variables for the Big Five factors of personality description that were identified by Ostendorf (1990). He translated the adjective rating scales employed by Norman (1963), McCrae and Costa (1985, 1987), Goldberg (1983; see also Goldberg, 1992), Peabody (1984, 1987), and Peabody and Goldberg (1989), and combined them with a list of German adjectives proposed by John (1982, see John, Goldberg and Angleitner, 1984).

To make the sample comparable with the Adult NEO-FFI Study, participants younger than 18 or older than 24 years were excluded. Subsequently, the sample was balanced for sex ($N = 568$, 284 males and 284 females; mean age 21.6 years, SD = 1.70 years) by random exclusion of an appropriate number of females.

The self-ratings on the 179 bipolar adjectives were factor analysed (principal components analysis followed by varimax rotation). Inspection of the scree plot suggested a five-factor solution that explained 43.3% of the variance. These factors corresponded well to the five-factor model of personality description as can be seen from the highest-loaded item on each factor: extraversion (sociable–unsociable), conscientiousness (careful–careless), neuroticism (tense–relaxed), agreeableness (peaceable–quarrelsome), and culture (intellectual–thoughtless). For each factor, a scale was constructed that consisted of the 12 items with the highest factor loadings and cross-loadings below 0.40. The internal consistencies of the five scales were high (in each case, $\alpha > 0.87$).

Berlin relationship study

Sample

When students of Humboldt University, Berlin, enrolled a few weeks before their first term opened, they were asked to participate in a longitudinal study on students' social relationships. Only students below 23 years of age were included. During the second week of their first term, 173 females (age 18–22 years, $M = 20.0$) and 64 males (age 18–22 years, $M = 20.4$) participated in the first session. They constituted the sample of the study by Asendorpf and Wilpers (1998). Because of the smaller male sample, we repeated the study one year later with a second sample of 75 males (age 19–24 years, $M = 20.8$). Because the results for the two male samples were virtually identical with regard to all

major variables, we pooled the two male samples, resulting in a more sex-balanced sample ($N = 312$; 173 females, 139 males).

Seven assessments over a period of 18 months were scheduled. The present analyses refer to the four assessments of personality every 6 months, and the first assessment of loneliness, self-esteem, relationship-specific attachment, and the social network; this assessment took place in the second week of the first term. A total of 171 students (92 females, 79 males) continued to participate until the last assessment. A volunteer subsample of 144 students (79 females, 65 males) maintained an additional diary for 21 days in the middle of their first term.

Personality, self-esteem, and loneliness

The Big Five factors of personality were assessed by the German version of the NEO-FFI by Costa and McCrae (1989, Borkenau and Ostendorf, 1993). Reliabilities and stabilities were highly similar to those reported by Asendorpf and Wilpers (1998). Global self-esteem as well as self-esteem toward same- and opposite-sex peers were assessed by German short versions of the SDQIII by Marsh and O'Neill (1984). The six items with the highest corrected item–scale correlations in the original questionnaire were selected and translated into German. The resulting six-item scales showed satisfactory reliabilities ($\alpha > 0.79$). Loneliness was assessed by a short version of the UCLA loneliness scale. A German version of the UCLA scale (Döring and Bortz, 1993) was reduced to the five highest-loading items on the first factor, representing feelings of loneliness, and the five highest-loading items on the second factor, representing feelings of social isolation, in a factor analysis of all 20 items in a representative sample of the general German population ($N = 592$). The resulting ten-item scale was highly reliable ($\alpha = 0.91$).

Security of attachment

Security of attachment to mother, father, same-sex peers, and opposite-sex peers were each measured by six-item scales. The items were identical across scales except that they referred to a different type of relationship. The items were selected on the basis of factor analysis from a larger item pool that was derived from the prototypic descriptions of the four attachment styles provided by Bartholomew and Horowitz (1991) (see Asendorpf and Wilpers (2000) for details). Each item was answered on a five-point Likert scale of an agreement format ('not at all true'–'completely true'). The reliabilities of the resulting six-item scales were satisfactory for mother and father ($\alpha > 0.82$) and acceptable for peers ($\alpha = 0.71$); see Asendorpf and Wilpers, 2000, for more information on these scales).

Relationship questionnaire

The participants listed all persons who were currently personally important to them, indicated their sex and age and the duration of the relationship with them, and rated the quality of the relationship during the last 3 months on eight Likert scales (see Asendorpf and Wilpers, 1998, for more details). In the present study, five scales produced significant findings: *contact frequency* (six-point scale 0–5, 'less than once a month'–'daily'), *closeness* of the relationship (five-point scale 1–5, 'very distant'–'very close'), *available support* (five-point scale 1–5, 'If I have problems, I would turn to this person to talk about my problems', 'never'–'always'), *frequency of conflict* (five-point scale 1–5, 'never'–'nearly at every encounter'), and *in love* (five-point scale 1–5, 'not at all'–'very much so').

Social interaction diary

The participants reported the beginning and end of any interaction that lasted for at least 10 minutes or was emotionally arousing for them, identified the interaction partner and rated

the interaction on nine Likert scales (see Asendorpf and Wilpers, 1998, for more details). The 144 participants reported an average of 9.5 interactions per day for 14–23 days ($M = 20.8$ days). Five variables produced significant findings in the present study: *number of interactions* with particular interaction partners per day, *percentage of interactions* spent with particular interaction partners, the *size of the group* in which the interaction took place (minimum two), and the ratings of *conflict* (five-point scales 1–5, 'no'–'very strong') and *perceived liking by the interaction partner* (seven-point scale 1–7, 1, 'dislikes me', 4, 'neither/nor', 7, 'likes me').

LOGIC study (child Q-sort study and child adjective study)

Sample

The LOGIC sample originally consisted of 204 children who started to attend 20 preschools in the Munich area in the autumn of 1984 when they were 3–4 years old, and whose first language was German; after 1 year, another 26 participants of the same birth cohort were added to the sample. This initial sample of 230 children (119 boys, 111 girls) was fairly unbiased because the schools were selected from a broad spectrum of neighbourhoods, and more than 90% of the parents whose permission was asked gave their consent for their child's participation. Attrition was relatively low and unsystematic (19% over 9 years, mostly due to a change in residence). However, not all parents and friends of all participating children wanted to cooperate in the study (see Weinert and Schneider, 1999, for more details).

Parental Q-sort (child Q-sort study)

The 54-item short version of the California Child Q-Set (Block and Block, 1980) was adapted to German by bilingual parents (Göttert and Asendorpf, 1989). This short form is representative of the full 100-item Q-sort with regard to its two major dimensions: ego-control and ego-resiliency (see Block and Block, 1980). At age 10, the child's main caregiver (nearly always the mother) received a letter with the Q-items, a detailed instruction for the Q-sort procedure, and the number of a telephone hotline for any questions about the procedure. A total of 170 parents provided a Q-sort description of their child. They sorted the 54 items for their judged salience in the child according to a forced equal nine-point distribution, ranging from 'extremely uncharacteristic' to 'extremely characteristic'. The resulting Q-sort pattern served as the person-centred parental description of the child.

Parental big five questionnaire (child adjective study)

At age 12, 155 participants were judged by their main caregiver (nearly always the mother) on bipolar adjectives that were obtained from the set of adjectives that were used in the Adult Adjective Study. The 24 highest-loading items on each factor of the adult self-ratings were pretested with ten German children from Grades 5 and 6 who were asked to mark those adjectives that they did not fully understand in terms of their meaning as personality descriptors. Each adjective marked by any one child was deleted from the list, and the highest-loading 12 bipolar items per factor were retained. The order of the adjectives within pairs was balanced with regard to the social desirability of the items. Finally, the order of the 60 bipolar items was randomized. Each bipolar item was answered on a five-point scale ('very–somewhat–neither/nor–somewhat–very').

Each *a priori* 12-item scale was reduced to an eight-item scale on the basis of the items' factor loadings and item–scale correlations (see Asendorpf and van Aken, 1999, for more details). The reliabilities for these five Big Five scales were satisfactory, median $\alpha = 0.86$,

range 0.83–0.91. The items with the highest corrected item–total correlation per scale were 'sociable–withdrawn', extraversion; ' helpless–self-assured', neuroticism; 'good–natured–touchy', agreeableness; 'thorough–careless', conscientiousness; 'knowledgeable–uneducated', openness. The resulting Big Five pattern served as the person-centred parental description of the child.

Shyness, sociability, aggressiveness, and loneliness scales

At age 12, all participants in the LOGIC study ($N = 186$) answered 15 items that referred to shyness (alternatively labelled inhibition by Asendorpf and van Aken, 1999), sociability, aggressiveness, and loneliness on seven-point Likert scales ('never'–'always'). The four items of the shyness scale were 'I am shy toward unfamiliar children', 'I easily approach unknown children' (reversed), 'When I meet unknown adults, I need a long time to warm up', 'I am somewhat inhibited toward unknown adults'. The four items of the sociability scale referred to a preference for being with people rather than being alone, and the four items of the aggressiveness scale to aggressiveness with peers. These two scales are identical with the scales used by Asendorpf and Meier (1993). The three loneliness items were 'I am lonely', 'I feel alone', and 'I feel left out of many things'. Thus, they refer to subjective feelings of loneliness, not to the number of friends or similar indicators of social integration. The reliabilities α were only marginal for two of the children's self-ratings (loneliness, 0.81; sociability, 0.78; aggressiveness, 0.66; shyness, 0.60).

The 155 parents, and 125 friends that were nominated by the children, answered the same questionnaire except that the loneliness items were dropped because it made no sense to have others rate one's subjective feelings. The items were adapted by replacing 'I' by 'my child' or 'my friend', respectively. The items of the parental shyness scale were identical with those used in earlier assessments (see, e.g. Asendorpf and van Aken, 1994). The reliabilities were satisfactory for the three parental scales ($\alpha > 0.85$), and acceptable for the three friend scales ($\alpha > 0.75$).

Cognitive and social self-esteem

At age 12, 183 participants were tested for cognitive and social self-esteem with a German version of Harter's Self-Perception Profile (Asendorpf and van Aken, 1993; Harter, 1985). The reliabilities were acceptable (for scholastic competence, $\alpha = 0.72$, for social acceptance, $\alpha = 0.78$). Because of a strong reference group effect for cognitive self-esteem due to different school tracks, the cognitive self-esteem scores were corrected for this effect (see Asendorpf and van Aken, 1994).

Derivation of personality prototypes

Prototypes of personality patterns were derived by the same method in the two adult studies of Big Five patterns and the LOGIC Adjective Study. In each case, individual patterns of raw Big Five scale scores were grouped into three, four, and five clusters through the two-step clustering procedure used by Caspi and Silva (1995). In contrast, the Q-sort patterns that were obtained in the LOGIC assessment at age 10 were grouped through the four-step procedure that was originally developed by Robins *et al.* (1996) and that was also used by Asendorpf and van Aken (1999).

Q-factor analyses of Q-sort patterns

The Q-sort patterns were first Q-factor analysed separately for three, four, and five factors. Second, the replicability of the factors was studied, and only replicable solutions were retained. Third, Q-sort patterns that loaded particularly highly on one factor were grouped

into prototypic Q-types. Fourth, the remaining patterns were assigned to these Q-types through discriminant analysis if the probability of correct classification of a pattern was high. Therefore only 146 of the 170 children (86%) could be classified. This classification method is described in more detail by Robins *et al.* (1996) and Asendorpf and van Aken (1999).

The three-, four-, and five-prototype solutions were compared within each study with regard to their replicability. Replicability was evaluated by comparing the solutions between random halves of each sample of participants. For each sample, ten different random splits within sex (and within age in the adult samples) were studied (using the SPSS command SAMPLE with different random number seeds). The clusters that were derived through cluster analysis were compared with a method originally proposed by Breckenridge (1989) that is described in detail in the section on cluster analysis. The Q-factors were compared with Pearson correlations of the Q-factor scores of the items (Robins *et al.*, 1996, and Asendorpf and van Aken, 1999, used the same method). The factors of one of the two random halves were reordered as to maximize the mean replicability across all factors. Q-factor solutions with a median replicability of at least 0.80 were considered acceptable.

Cluster analyses of big five patterns

First, Ward's hierarchical clustering procedure was applied. The two most similar clusters in terms of their squared Euclidean distance were combined stepwise, starting with the clusters that contained only one participant. The three-, four-, and five-cluster solutions were considered for further analysis. Because clusters that are fused in one step remain together in all later steps, this procedure can result in nonoptimal solutions. Therefore, each solution was subsequently used as the initial cluster centres for a nonhierarchical *K*-means clustering procedure (SPSS procedure QUICK CLUSTER, option NOUPDATE). In each analysis, all participants were assigned to the most similar cluster on the basis of their Euclidean distance from the initial cluster centres. Subsequently, new cluster centres were computed and used as new initial cluster centres for the next step in an iterative procedure until the largest change in any cluster centre was less than 2% of the minimum distance between the initial centres (the SPSS default criterion).

To evaluate the replicability of the final solutions, the full sample of participants was randomly split into halves as described in the 'Overview' Section. The full two-step procedure (Ward, followed by *k*-means) was applied to each half. The two solutions were then compared for agreement as follows. The participants of each half of the sample were assigned to new clusters on the basis of their Euclidean distances to the cluster centers of the other half of the sample (SPSS procedure QUICK CLUSTER, option CLASSIFY). These new clusters were then compared for agreement with the original clusters with Cohen's κ (sometimes it was necessary to reorder the clusters to increase agreement; the order of the clusters is irrelevant). The two resulting κs were averaged. As in other applications of Cohen's κ, an agreement of at least 0.60 was considered acceptable.

RESULTS

Replicability of the prototypes

The median replicability of the three-, four-, and five-prototype solution in the four studies is presented in Table 1. For the cluster analyses, replicability was acceptable only for the

Table 1. Replicability of personality prototypes in four studies

Study	Judge	*N*	Median replicability for prototypes		
			3	4	5
Adult NEO-FFI	Self	730	$\kappa = 0.69$	$\kappa = 0.54$	$\kappa = 0.51$
Adult adjectives	Self	568	$\kappa = 0.66$	$\kappa = 0.55$	$\kappa = 0.49$
Child adjectives	Parent	155	$\kappa = 0.60$	$\kappa = 0.43$	$\kappa = 0.37$
Child Q-sort	Parent	170	$r = 0.85$	$r = 0.66$	$r = 0.64$

Median replicability refers to ten random splits of the sample within sex (and within age for the adult studies).
The κ replicabilities can be compared with each other but not with the *r* replicabilities.
The *r* replicabilities are medians across factors and random splits.

three-cluster solutions. For the Q-factors, all three factors of the three-factor solution were sufficiently replicable (factor 1, 0.92; factor 2, 0.83; factor 3, 0.81). For the four-factor and the five-factor solutions, only the first factor was replicable (in both cases, $r > 0.84$; for all other factors, $r < 0.63$). Thus, both with regard to the overall replicability and with regard to the replicability of specific factors, only the three-factor solution was acceptable.

The replicability coefficients varied considerably across the ten random splits within each study. For the three-prototype solution, they ranged from 0.40 to 0.83 (Adult NEO-FFI Study), from 0.35 to 0.79 (Adult Adjective Study), from 0.37 to 0.84 (Child Adjective Study), and from 0.64 to 0.87 (Child Q-Sort Study). Therefore it was important to evaluate replicability with multiple random splits for each sample of participants. Because replicability was only just acceptable for the Child Adjective Study, an additional replicability analysis was performed with 50 random splits. Replicability varied between 0.31 and 0.89 with a median of 0.61. Thus, ten random splits produced stable median results even for a relatively small sample of participants. All in all, these analyses suggest that only three personality prototypes were replicable, independent of age, judge, and method of analysis. Therefore, only the three-prototype solutions were further analysed.

Description of the three prototypes

Figure 2 describes the three prototypes that resulted in the three studies of Big Five patterns in terms of the mean Big Five pattern of the members of the corresponding cluster (panels (A)–(C)). To describe the Q-types in a comparable format, the Big Five scores of the members of a type were averaged (panel (D)). Such scores were available for 122 of the 146 Q-classified children. To facilitate comparison of these patterns across studies, they were based on sex-balanced samples (three boys were randomly dropped at age 12; the 122 children at age 10 were accidentally balanced for sex), and standardized (*z*-transformation for the full sex-balanced samples). Because the Q-types were based on Q-sorts obtained at age 10, but described by Big Five patterns at age 12, the Big Five patterns of the Q-types have a smaller scatter (less extreme *z*-scores).

Figure 2 shows that in each study one of the patterns could be unambiguously identified as resilient (low neuroticism, above-average scores otherwise). A minor deviation occurred only for the NEO-FFI study (panel (A)), where the openness scores in this cluster were slightly below average. Furthermore, one pattern in both adult studies could be clearly identified as overcontrolled, and one pattern in both child studies as under-controlled. However, the overcontrolled pattern in the Adult Adjective Study was

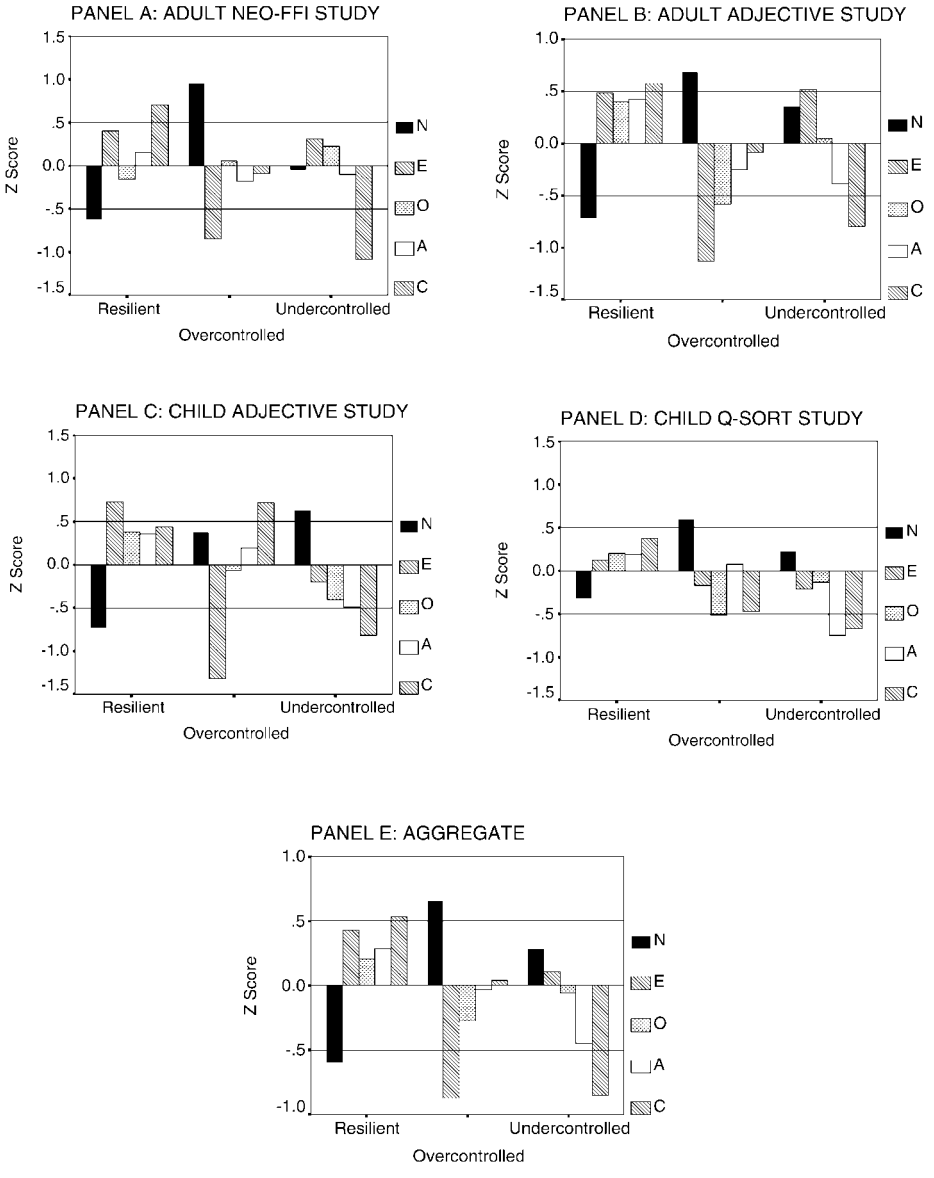

Figure 2. Three major personality prototypes characterized by their Big Five patterns in four studies and across the four studies.

characterized by unexpectedly low openness, and the undercontrolled pattern in the Child Adjective Study by unexpectedly high neuroticism (see Figure 2).

The remaining patterns could be identified as undercontrolled in the adult studies, and as overcontrolled in the child studies, but only with reservations. The undercontrolled pattern based on the NEO-FFI was characterized by average rather than low agreeableness, and the overcontrolled patterns in the LOGIC study by unexpectedly high conscientiousness (Adjective Study) and average rather than low extraversion and unexpectedly low openness (Q-Sort Study; see Figure 2).

It is important to note that these deviations from the expected patterns were not consistent between the two adult studies (panels (A) versus (B) in Figure 2), nor between the two child studies ((C) versus (D)), nor between the two studies using bipolar adjectives ((B) versus (C)). Thus, they reflected specifics of the studies, not systematic age, judge, sample, or method effects. In such instances, aggregating the results across studies is a useful method for carving out consistencies. Panel (E) shows the mean pattern across the four studies for the resilient, the overcontrolled, and the undercontrolled prototype. These patterns fitted the expected patterns very well. Resilients had socially desirable scores in all five scales, overcontrollers were introverted and neurotic, and undercontrollers were disagreeable and unconscientious. Thus, despite study-specific deviations from what was expected, the overall findings strongly confirmed the hypothesis.

Inspection of the group sizes for the three prototypes (i.e. the cluster or the Q-type sizes) confirmed the expectation that the resilient prototype characterized the largest type. The proportion of the overcontrolled and the undercontrolled groups varied across studies with no consistency between the two adult, the two child, or the two adjective studies. Therefore, a useful estimate of the relative frequencies of the three personality patterns was their mean proportion across the four studies. These mean proportions were 49% resilient, 23% overcontrolled, and 28% undercontrolled patterns.

Sex differences in the frequencies of the three personality patterns were tested by χ^2-tests. The overall tests were not even marginal for the adult studies, and marginal for the child studies (in both cases, $\chi^2(\mathrm{df}=2)>4.7$, $p>0.10$). Because of the relatively small child samples, *post hoc* tests explored specific group differences. The sex proportion for each pattern was compared with the sex proportion for the two pooled remaining patterns. The results were consistent across the two studies. Boys were more often judged as undercontrolled than girls, Adjective Study, $\chi^2(\mathrm{df}=1, n=155)=6.77$, $p>0.01$, 39% undercontrolled boys, 19% undercontrolled girls, Q-Sort Study, $\chi^2(\mathrm{df}=1, n=146)=3.23$, $p>0.08$, 26% undercontrolled boys, 14% undercontrolled girls; and girls were more often judged as resilient than boys, Adjective Study, $\chi^2(\mathrm{df}=1, n=155)=4.65$, $p>0.05$, 61% girls, 44% boys, Q-Sort Study, $\chi^2(\mathrm{df}=1, n=146)=4.62$, $p>0.04$, 70% girls, 53% boys. Thus, independent of the method of deriving the prototypes, boys were more often classified as undercontrolled than girls, and less often classified as resilient, whereas no sex differences were found for adulthood.

Consistency of the prototypes across studies

Aggregation across studies is a rough method of revealing cross-study consistencies. Better suited is a quantitative index of consistency. For the three studies of Big Five patterns, the consistency of the cluster centres was evaluated with the same method that was used for the replicability within studies. To make the Big Five patterns comparable across different studies, both the individual patterns and the cluster centres were standardized within each sex-balanced sample (z-transformation). Subsequently the participants of one study were assigned to the best-fitting cluster of another study using the SPSS procedure QUICK CLUSTER, option CLASSIFY, and this new classification was cross-classified with the original classification. Because this procedure could also be applied to the other study, the two resulting κs were averaged. These consistency coefficients can be directly compared to the replicability coefficients (see Table 2).

Two main observations can be made from Table 2. First, the consistency between the two studies with adults was as high as the within-study replicabilities. This result can be

Table 2. Consistency of the prototypes across different studies

Study	Judge	*N*	Adult		Child	
			NEO-FFI	Adjectives	Adjectives	Q-sort
Adult NEO-FFI	Self	730	0.69	0.70	0.51	0.46
Adult adjectives	Self	568		0.64	0.64	0.54
Child adjectives	Parent	152			0.60	0.52
Child Q-sort	Parent	122				–

All samples were balanced with regard to sex.
Reported are κ coefficients.
Replicability coefficients within studies are reported in italics.
No κ replicability could be computed for the Q-sort data.

attributed to the fact that the replicability coefficients were based on only half of the sample whereas the consistency coefficients were based on the full samples and therefore referred to more reliable clusters. It seems that this increase in cluster reliability had a similar effect on the cross-study consistency as the consistency-decreasing effect of the different method of Big Five assessment (NEO-FFI versus adjectives).

Second, the consistency between the two adult studies was higher than their consistency with the two child studies. This difference can be attributed not only to the different age of the participants but also to the fact that both adult studies used self-ratings whereas both child studies used parental ratings (thus, a judge effect).

The consistency between the Q-sort prototypes and the other prototypes was computed by using the Big Five means of the Q-types (see Figure 2(D)) as the prototypic patterns of the Q-prototypes. This method underestimates somewhat the consistency between the age 10 Q-prototypes and the other prototypes because (a) the Q-prototypes were described with Big Five patterns that were assessed two years later and (b) these patterns were based on only 122 of the 146 Q-typed children because some of the children dropped out from the study and some parents did not cooperate at the later assessment.

Table 2 indicates that the consistency between the Q-prototypes and the adult Big Five prototypes was similar to the consistency between the child Big Five prototypes and the adult Big Five prototypes. At a first glance, this result might suggest a high consistency between the Q-typing and the clustering method, but Table 2 shows that the consistency between these two methods was not higher than the consistency between the Q-prototypes and the adult prototypes. This relatively low consistency between the two methods for similar samples of children and judges may be attributed to the much smaller sample of cross-classified participants (152 versus 568 and 730) that may have increased classification errors.

All in all, the replicability and consistency analyses show that the three personality prototypes are not only replicable within the same sample of participants and judges but are also similarly consistent across different ages, judges, and methods of deriving the prototypes.

Relations between the individual prototypicality scores

To test the hypothesis of a quadratic relationship between resiliency and overcontrol/undercontrol, individual prototypicality scores for the three personality prototypes were computed. As in the Asendorpf-e and van Aken (1999) study, children's Q-factor loadings

on the three Q-factors served as prototypicality scores in the Q-Sort Study. A child's loading on a Q-factor is identical with the Pearson correlation between the child's Q-sort pattern and the pattern of the Q-factor scores of the Q-factor. Similarly, the Euclidean distance between an individual Big Five pattern and a prototype's Big Five pattern was considered a measure of this individual's nonprototypicality for the prototype, and unity minus this distance a measure of the individual's prototypicality. Thus, if an individual pattern corresponded perfectly with a prototype, the Euclidean distance was zero and the prototypicality score was unity; lower prototypicalities were indicated by larger distances and hence lower prototypicality scores. Intraindividual correlations cannot be used in this case because the mean level of the individual patterns can vary (in contrast, it is constant for California Child Q-Sort patterns because of the forced equal intraindividual distribution of these patterns).

As in the Asendorpf-e van Aken (1999) study, quadratic relations between the three prototypes were evaluated by regressing a prototypicality score first on another prototypicality score and then on its square in a hierarchical regression analysis. The expected quadratic relation between resiliency and the other two prototypes was significantly confirmed in all four studies (in each of the eight cases, $F > 7.6$, $p > 0.01$, $R^2_{\text{change}} > 0.029$), and the expected lack of a quadratic relation between overcontrol and undercontrol was also confirmed in three of the four studies (in each of the three cases, $F < 2.2$, $p > 0.14$, $R^2_{\text{change}} < 0.013$). The only significant quadratic relation between overcontrol and undercontrol was found in the NEO-FFI Study, $F(1, 727) = 17.36$, $p < 0.001$, $R^2_{\text{change}} = 0.017$) but it was relatively weak. Thus, the hypothesis of quadratic relations between resiliency and the other two prototypes was fully confirmed.

Bimodality of the prototypicality distributions

As in the Asendorpf-e van Aken (1999) study, for each pair of types (clusters or Q-types), the third type was dropped from analysis, and the remaining distribution of the prototypicalities of the two types was plotted, dividing the prototypicality continuum into ten equal intervals. Thus, 4 (study) $\times$ 3 (pairs of types) $\times$ 2 (prototypicalities) $= 24$ distributions were studied. The expected bimodal patterns for overcontrol and undercontrol when the resilient type was dropped from analysis was found only for the Q-sort data. In all other 22 cases, the distribution was unimodal. Thus, the bimodality hypothesis could not be replicated for instruments other instruments other than for the California Child Q-Set.

Reasons for this failure of generalization may be an uneven sampling of the items of (a) the full 100-item California Child Q-Set or (b) our short 54-item version of it. In order to decide between these two explanations, we analysed a Dutch sample of 170 4- to 7-year -old children who were described on the full 100-item Q-sort by their teachers (we thank Gerbert Haselager for providing these data to us). Again, these data showed quadratic relations between the prototypicalities for resiliency, and overcontrol and undercontrol, but no bimodal distribution for over- or undercontrol when resilients were dropped from analysis. Thus, the mixed discrete–fuzzy model is specific to the short 54-item Q-sort, which may be due to an uneven selection of items for the short version.

Stability of classification in adulthood

The stability of the classification for the prototypes was studied across the four assessments of the NEO-FFI in the Berlin Relationship Study. Participants were not

Table 3. Stability of participants' classifications for the three prototypes in the Berlin relationship study

Assessment	Assessment		
	After 6 months	After 12 months	After 18 months
Initial	0.61	0.47	0.54
After 6 months		0.63	0.58
After 12 months			0.61

$N = 171$.
Reported are κ coefficients.

classified for cluster solutions that were derived separately for each assessment. Instead, in order to increase the reliability of the classification, the 171 participants of the longitudinal sample were classified separately for each assessment with regard to the prototypes that were derived from the larger NEO-FFI Study ($N = 730$, see Figure 2(A)) using the SPSS procedure QUICK CLUSTER, option CLASSIFY. Subsequently, these four classifications for each participant were cross-classified using Cohen's κ as a measure of stability.

Table 3 indicates that the 6 month stabilities were similar to the replicability of the prototypes (see Table 2) and decreased slightly over longer retest intervals. Thus, the 6 month stabilities showed a sufficient retest reliability of the prototypes.

Correlates of the prototypes in adulthood

Two different strategies of assessing social correlates of the three prototypes in the Berlin Relationship Study are possible. A first strategy capitalizes on aggregation by classifying the participants on the basis of their aggregated NEO-FFI patterns from all four assessments. However, this strategy leads to biased results because undercontrollers were underrepresented in the longitudinal sample. This became obvious when the 312 participants in the first assessment were classified with regard to the prototypes that were derived from the larger NEO-FFI Study (SPSS procedure QUICK CLUSTER, option CLASSIFY). The proportions of the three resulting clusters were virtually identical with the proportions in the NEO-FFI Study. However, when these proportions were compared between the longitudinal sample ($n = 171$) and the drop-outs ($n = 141$), a significant difference was found, $\chi^2(\mathrm{df} = 1,\ n = 312) = 13.46,\ p > 0.001$. Undercontrollers were relatively less frequent in the longitudinal sample than in the drop-out sample (19% versus 35%), and resilients were relatively more frequent (52% versus 34%). In contrast, the proportions of the overcontrollers were highly similar. Thus, differences between undercontrollers and resilients were attenuated in the longitudinal sample by a systematic drop-out effect.

Therefore we studied correlates of the classification of all 312 participants in the first assessment. Because of the large number of correlates, it was necessary to minimize chance findings. We used a moderately conservative procedure of testing for cluster differences. For each variable, overall cluster differences were tested by an ANOVA with a significance level of $p = 0.01$ except for the much smaller diary sample where we set the significance level to $p = 0.02$ (thus, similar effect sizes had a similar chance to become significant in all cases). If overall differences were found, we tested differences between the three clusters by Student–Newman–Keuls *post hoc* tests.

Table 4. Significant cluster differences in personality in the Berlin Relationship Study

Variable	Cluster		
	Resilient	Undercontrolled	Overcontrolled
Shyness	2.58[a] (0.78)	2.59[a] (0.78)	3.19[b] (0.81)
Sociability	3.57[a] (0.64)	3.69[a] (0.64)	3.10[b] (0.66)
Loneliness	1.69[a] (0.61)	1.72[a] (0.62)	2.36[b] (0.66)
Self-esteem			
global	4.12[a] (0.51)	3.77[b] (0.63)	3.15[c] (0.80)
with same-sex peers	3.81[a] (0.57)	3.82[a] (0.56)	3.47[b] (0.60)
with opposite-sex peers	3.62[a] (0.79)	3.66[a] (0.76)	3.27[b] (0.81)

$N = 312$.

Reported are cluster means (standard deviations in parentheses).

Means with the same superscript are not significantly different according to Student–Newman–Keuls tests ($p < 0.05$).

Personality correlates

Because the scales assessed internalizing difficulties rather than externalizing difficulties, overcontrollers were expected to differ on all scales from the other participants whereas no differences between resilients and undercontrollers were expected. As Table 4 indicates, this pattern was confirmed with only one exception: undercontrollers reported a lower global self-esteem than resilients. Because they reported an equally high self-esteem with peers, this difference may reflect problems with parents. Below we report results that support this interpretation.

Social relationship correlates

Significant differences in the relationship questionnaire and the attachment scales are presented in Table 5. Concerning structural features of their social network, overcontrollers reported fewer relationships than both undercontrollers and resilients (the difference with resilients was significant only for peers), whereas undercontrollers had relationships with younger people and of a shorter duration than both overcontrollers and resilients.

Concerning the quality of relationships, undercontrollers reported more overall interpersonal conflict, a less close and more insecure relationship with their mother, and more satisfaction with the frequency of their peer contacts than resilients (see Table 5). Analyses of the conflict ratings for specific relationships revealed that undercontrollers reported significantly more conflict with the mother, the peers, and the romantic partner (if they had one); for the father and siblings, similar tendencies did not reach significance. Thus, as expected, undercontrollers showed a generalized tendency to engage in interpersonal conflict. Interestingly, they perceived their relationship with their mother, but not with their peers, as distant and insecure.

In contrast, the overcontrollers reported insecure relationships with both mother and peers, and less available support from their father. Also, they interacted less often with the peers in their network and fell less often in love with them. Contrary to expectation, the overcontrollers perceived more conflict in their relationships than the resilient participants. Analyses of the conflict ratings for specific relationships revealed that they reported significantly more conflict with peers but not significantly more conflict with their mother, father, siblings, or romantic partner. Thus, their higher conflict ratings were less generalized than the conflict ratings of the undercontrollers.

Table 5. Significant cluster differences in social relationships in the Berlin Relationship Study

Variable	Cluster		
	Resilient	Undercontrolled	Overcontrolled
Number of relationships	25.7[ab] (10.1)	28.3[a] (10.1)	23.6[b] (7.89)
with peers (age 18–27)	15.0[a] (6.75)	16.6[a] (7.61)	13.0[b] (5.49)
Age of relationship partners (years)	29.5[a] (3.95)	27.8[b] (3.98)	29.5[a] (4.53)
Duration of relationships (years)	8.10[a] (2.09)	7.01[b] (2.36)	8.00[a] (2.51)
Frequency of conflict	1.69[a] (0.46)	1.89[b] (0.44)	1.94[b] (0.46)
Closeness to mother	4.36[a] (0.94)	3.91[b] (0.99)	4.16[ab] (1.12)
Available support from father	3.10[a] (1.23)	2.99[a] (1.22)	2.59[b] (1.05)
Contact frequency with peers	2.26[a] (0.74)	2.39[a] (0.63)	2.01[b] (0.64)
In love with peers	1.78[a] (1.75)	1.49[ab] (1.80)	1.18[b] (1.66)
Security of attachment to			
mother	4.18[a] (0.75)	3.75[b] (0.80)	3.78[b] (0.95)
same-sex friends	3.80[a] (0.56)	3.72[a] (0.58)	3.35[b] (0.57)
opposite-sex friends	3.63[a] (0.57)	3.62[a] (0.62)	3.18[b] (0.61)

$N = 312$ (308 for relationship with mother, 293 for relationship with father).
Reported are means (standard deviations in parentheses) across cluster members and relationships of a particular type.
Means with the same superscript are not significantly different according to Student–Newman–Keuls tests ($p < 0.05$).

It is important to note that the peer differences cannot be explained by the different living conditions of these young adults. Resilients lived more frequently at home together with their parent(s) (49%) than both undercontrollers (35%) and overcontrollers (28%), $\chi^2(\mathrm{df} = 2, n = 312) = 11.42$, $p > 0.01$, but overcontrollers reported more problematic peer relationships than both undercontrollers and resilients. Similarly, the cluster differences with regard to the parental relationships did not correspond to the cluster differences in the living condition. Instead, the cluster differences seem to be due to personality traits of the participants.

Social interaction correlates

The three clusters did not differ with regard to most global characteristics of social interaction such as number of interactions per day, percentage of time spent with interaction, or interaction qualities that were averaged across all interactions. As Table 6 indicates, undercontrollers interacted with younger people than the other participants (the difference was significant only for overcontrollers), replicating the similar effect for the relationship questionnaire. Overcontrollers interacted in smaller groups than resilients. Closer inspection revealed that they were involved in as many dyadic interactions as the other participants but that they interacted significantly less in groups consisting of three or more people, $t(142) = 2.40$, $p > 0.02$. Thus, they tended to avoid group interactions.

The remaining cluster differences referred specifically to interactions with the mother and with peers. Undercontrollers reported more conflict in their (relatively rare) interactions with their mother than the other participants. Overcontrollers had relatively many interactions with their mother, fewer peer interactions than the resilients, and felt more than both resilients and undercontrollers that their peers did not like them.

Again, the parent versus peer differences cannot be attributed to the living conditions of the participants. Similar to the full sample, the resilient diary participants tended to live

Table 6. Significant cluster differences in social interaction in the Berlin Relationship Study

Variable	Cluster		
	Resilient	Undercontrolled	Overcontrolled
Age of interaction partners (years)	27.6^{ab} (4.78)	25.6^{a} (3.65)	28.9^{b} (5.42)
Group size for interactions	3.80^{a} (0.99)	3.66^{ab} (0.87)	3.32^{b} (0.63)
Percent interactions with mother	9.30^{ab} (8.41)	6.28^{a} (4.82)	11.18^{b} (8.60)
Conflict with mother	1.26^{a} (0.31)	1.61^{b} (0.85)	1.28^{a} (0.35)
Number of peer interactions/day	6.05^{a} (3.20)	5.41^{ab} (2.94)	4.46^{b} (2.57)
Perceived liking by peers	5.11^{a} (0.47)	5.16^{a} (0.44)	4.85^{b} (0.73)

$N = 144$ (142 for interactions with mother).
Reported are means (standard deviations in parentheses) across cluster members and interactions of a particular type.
Means with the same superscript are not significantly different according to Student–Newman–Keuls tests ($p < 0.05$).

more frequently at home (49%) than both the undercontrolled participants (38%) and the overcontrolled participants (33%), but the difference was not significant, and the overcontrollers had the highest rate of interaction with the mother and a low frequency of peer interactions, but the lowest rate of living at home.

Correlates of the prototypes in childhood

Concurrent personality correlates of the three prototypes in childhood that correspond to the personality correlates in adulthood are reported here for the prototypes that were derived from the Big Five patterns in the LOGIC study (see Asendorpf and van Aken, 1999, for correlates of the three prototypes that were based on Q-sort patterns). A study of developmental antecedents of the three prototypes is possible with the LOGIC data but is beyond the scope of this article. We used the same procedure of testing for cluster differences as for the Berlin Relationship Study, setting the significance level for the ANOVAs to $p = 0.02$ because of the relatively small sample.

Table 7 shows the same internalizing pattern for overcontrollers that was also found for adults (see Table 5): more shy, less sociable, more lonely, and lower in social self-esteem than the resilients. As expected, undercontrollers were rated higher in aggressiveness by their parents than the other participants; for the self- and friend ratings, the same pattern was found but did not reach significance.

Two unexpected results were that parents (but not friends) judged undercontrollers lower in sociability than resilients (but the ratings were still significantly higher than for overcontrollers), and that undercontrollers reported a lower social self-esteem than resilients (which was however still significantly higher than overcontrollers' social self-esteem). In both cases, undercontrollers ranked between resilients and overcontrollers rather than together with resilients. Thus, overcontrollers showed strong internalizing tendencies whereas undercontrollers showed moderately strong externalizing and internalizing tendencies. This result is consistent with the unexpectedly high neuroticism of the undercontrollers in this study (see Figure 2(C)) and seems to be due to specifics of this study that cannot be generalized.

Table 7. Significant cluster differences in personality in the LOGIC study

Variable	Cluster		
	Resilient	Undercontrolled	Overcontrolled
Parent ratings			
shyness	2.65[a] (1.01)	3.02[a] (1.02)	3.98[b] (1.03)
sociability	5.69[a] (0.77)	5.08[b] (0.99)	4.48[c] (0.97)
aggressiveness	2.39[a] (0.73)	2.80[b] (1.01)	2.18[a] (0.79)
Friend ratings			
shyness	2.91[a] (1.06)	3.21[ab](1.05)	3.58[b] (1.29)
sociability	5.83[a] (0.87)	5.76[a] (0.81)	5.15[b] (1.16)
Self-ratings			
loneliness	1.88[a] (0.95)	2.16[ab] (0.94)	2.55[b] (1.18)
social self-esteem	3.15[a] (0.42)	2.93[b] (0.54)	2.78[b] (0.61)

$N = 155$ (125 for friends).
Reported are cluster means (standard deviations in parentheses).
Means with the same superscript are not significantly different according to Student–Newman–Keuls tests ($p < 0.05$).

DISCUSSION

The results of the present study (a) confirmed Caspi's conjecture of three generalizable personality types for both childhood and adulthood, (b) confirmed our stronger hypothesis that these and only these three types are replicable in large, unselected samples of persons and personality descriptors, (c) showed a sufficient 6 month retest stability of the types in adulthood, but a lower long-term stability, (d) showed a substantial consistency of the types across ages, judges, and methods of personality description, (e) confirmed that resilients are well adjusted, overcontrollers show internalizing tendencies, and undercontrollers show externalizing tendencies, (f) confirmed quadratic relationships between resiliency and both over- and undercontrol in all studies, and (g) failed to confirm a mixed continuous–categorical model for studies other than the Child Q-Sort Study. Together, these findings strongly support a fuzzy type model of personality description. Thus, there are reliable differences between personality patterns that are not captured by linear or curvilinear trait models.

The three-prototype model builds on two main traditions of personality assessment that have coexisted for a long time as fairly separate strands of research, the person-centred Q-sort approach and the variable-centred trait approach, and integrates them from a person-centred perspective. Thus, the present study can be seen as part of an increasing movement in personality psychology to bridge these two traditions (see e.g. John *et al.*, 1994; Robins *et al.*, 1996; van Lieshout *et al.*, 1995).

Confirmation of caspi's conjecture and our main hypothesis

Confirming Caspi's (1998) conjecture, there were three personality types at the coarsest level of differentiation, and they showed a substantial consistency across all four studies. Confirming our stronger hypothesis, these and only these three types were replicable. Note that Robins *et al.* (1996) and Hart *et al.* (1997) also found that only these three types were replicable. That the same number of replicable types was found in six independent studies despite the strong variation of the studies in terms of age (childhood, adolescence,

adulthood), language (English, German, Dutch, Icelandic), rater (self, teacher, parent, expert), method of personality description (NEO-FFI, bipolar adjectives, Q-sort), and method of deriving types is surprising, given the endless debate about the 'right' number of fundamental factors of personality traits.

One reason for the perfect consistency in the number of replicable types may be that the same criterion was used in all six studies: the replicability of the types between random halves of the sample. The replicability can be specified, as in the present study, by requiring median correlations between Q-factors of at least 0.80 for all factors, and median cluster replicabilities of at least 0.30 for all clusters, where the medians are computed for ten random splits of the sample. This specification provides a clear, absolute criterion that can be compared between studies.

Four notes of caution are in order at this place. First, it is not sufficient to rely on one random split of the sample because the results can vary considerably between different random splits. For example, when the sample of the Adult Adjective Study was split into participants with odd versus even codes, an acceptable replicability resulted for a four-type solution, and an unacceptable replicability for the three-type solution. The present analyses suggest that averaging the results of ten different random splits is necessary and sufficient for stable replicability results that no longer depend on the specific splits. Second, consistent results for the number of types can only be expected for large, unselected samples of persons and personality traits.

Third, if cluster analysis is used, the similarity of the profiles of the scale scores must be measured adequately (e.g., by the Euclidean distance, or its square). Measuring profile similarity by the Pearson correlation may produce misleading results because it ignores differences in the level and scatter of the profiles. Last but not least, it is important that our cluster analyses were always based upon raw scale mean scores. Transforming the scale raw scores into z scores (which equalizes the scale means and variances) or even factor scores (that additionally eliminate all correlations among the scales) leads, in our experience, to different and less replicable clusters. Whereas these transformations are not problematic, and one sometimes even desirable, in variable-centred analyses, they seriously distort the results of cluster analyses. Also, when the scales contain different numbers of items (e.g. the NEO-PI), analysing the scale sums is misleading because more weight is given to scales with more items. Scale means should be used in this case.

Whereas the number of the prototypes was identical across all four studies, their consistency was moderate to high (kappas ranging from 0.46 to 0.77). When the prototypic Big Five profiles of the three types were contrasted between any two studies, clear inconsistencies were found for at least two of the Big Five scales. However, these inconsistencies seem to be due to specifics of the studies, not to systematic factors such as age, type of judge, method of personality description, or method of deriving the types. When the prototypic profiles were averaged across the four studies, the resulting pattern confirmed very well with our *a priori* hypothesis that was based on the findings of Robins *et al.* (1996) and Caspi's (1998) informal review. Resilients had moderately socially desirable scores in all five scales; overcontrollers were introverted and neurotic; and undercontrollers were unconscientious and disagreeable; last but not least, in the remaining six cases (three for overcontrollers and three for undercontrollers), all scale scores deviated by less than 0.3 standard deviations from the sample means. Thus, overall, the resulting Big Five patterns confirmed well to the assumption that resilients are well adjusted, overcontrollers are introverted and neurotic, and undercontrollers are neither conscientious nor agreeable.

Short-term stability of the types in adulthood

The 6 month stability of the individual membership for the three types (averaged over the three retest intervals) was $\kappa = 0.62$ for the clusters based on the NEO-FFI and hence not much lower than the consistency of $\kappa = 0.69$ between random halves of the Adult NEO-FFI Study participants. Thus, the types were sufficiently stable over short periods of time. Note that these stability coefficients are Cohen's kappas that cannot be directly compared with the Pearson correlations that are used to measure the stability of traits and that typically range around 0.80 for 6 month intervals (see e.g. Asendorpf and Wilpers, 1998; McCrae and Costa, 1990).

Continuity and stability of the types between childhood and adulthood

The consistency of the prototypes between the adult and child studies implies a continuity of the three prototypes between late childhood and adulthood. In addition, Asendorpf and van Aken (1999) found a high continuity between early and late childhood: the Q-factor scores of teacher Q-sorts at ages 4–6 correlated between 0.88 and 0.78 with the Q-factor scores of parental Q-sorts at age 10. Note that these continuity data refer to the constancy of the prototypes, not to the stability of the individual membership for the three types. The stability can be low despite a high continuity. In fact, Asendorpf and van Aken (1999) found a significant but low stability of $\kappa = 0.30$ between ages 4–6 and 10 in the LOGIC study (which may underestimate the true stability because teachers served as judges at ages 4–6 and parents at age 10). Therefore we expect that future analyses of the stability of the types (e.g. in the Dunedin Longitudinal Study, Caspi and Silva, 1995) will find significant but only low stabilities between childhood and adulthood. Such low stabilities would not question the person-centred type approach because the stabilities in variable-centred analyses are also significant but low between childhood and adulthood for most traits (see e.g. Caspi, 1998). Instead, on the basis of the sufficient retest reliabilities, the low stabilities would indicate that the person-centred approach is well suited for developmental analyses of personality change.

Correlates of the three types

All in all, the hypothesis that overcontrollers show internalizing tendencies, and under-controllers show externalizing difficulties, was strongly confirmed. When overcontrollers were contrasted with resilients, they were more shy, less sociable, more lonely, and reported a lower self-esteem, both globally and in the social domain. These differences were found both for self-ratings of students and for parental, friend, and self-ratings of 12-year-old children. In contrast, when undercontrollers were contrasted with resilients, they were judged as more aggressive by their parents. Only two of the possible 28 comparisons in Tables 4 and 7 were unexpected. Undercontrollers were judged as less sociable by their parents than resilients (but as more sociable than overcontrollers, and this difference was not found for friend or self-ratings); and adult undercontrollers reported a lower global self-esteem than resilients (but a higher global self-esteem than over-controllers, and this difference was not found for the social self-esteem).

Analyses of the social relationships and daily interactions of the three types in adulthood also confirmed the internalizing/externalizing hypothesis. Overcontrollers reported fewer peer relationships, less contact with peers, fewer love relationships with peers, less secure relationships with peers of either sex, mother, and father, and less available support from their father than resilients. In a 3 week diary, they reported fewer group interactions, fewer

peer interactions per day, and thought that they were less liked by their peers than resilients. Thus, they reported problems of social acceptance and close relationships with peers but also with their parents. In contrast, undercontrollers reported relationships of shorter duration and with younger people, more interpersonal conflict across virtually all types of relationship, more often too much contact with peers, and a less close and secure relationship with their mother than resilients. In the diary they reported fewer interactions with their mother and more conflict in these interactions than resilients. Only two of the 38 comparisons were unexpected. Compared to resilients, overcontrollers reported more overall conflict in relationships (but closer analyses showed that this was only due to conflict with peers). Also, undercontrollers did not report more interpersonal conflict with peers in their diary than resilients.

Distribution of the personality patterns

When the individual prototypicality scores for the three types were analysed, the expected inverse U-shaped quadratic relations between the prototypicality for resiliency and the prototypicalities for overcontrol and undercontrol were confirmed in all studies. Thus, this finding by Asendorpf nd van Aken (1999) could be generalized. Contrary to expectation, the mixed discrete–fuzzy model of Asendorpf and van Aken (1999) (see Figure 1(D)) could not be generalized beyond the Child Q-Sort Study, and additional analyses showed that it was even specific to the short 54-item version of the California Child Q-Set.

Thus, what is generalizable is a model of personality patterns that confirms to Figure 1(C). There are three replicable, fuzzy personality types. The fact that the three prototypes were replicable within each study shows that a type approach could describe the data well, and the fact that the prototypes were generalizable across ages and methods indicates that they describe perceived personality differences, not specific assessment instruments. It is important to note that the existence of replicable types violates the multinormal linear model. Thus, the multinormal linear model does not sufficiently describes the distribution of personality patterns.

That the transitions between the types turned out to be smooth rather than discontinuous tells us that personality types at this highest level of analysis should be conceived as groups of people with fuzzy borders that can be distinguished from one another but that are not as discrete as lay concepts of personality types assume. That uninformed users of the term ‘personality type’ overestimate differences between the types does not imply, however, that the types are not useful scientific concepts

Limitations and problems of the present approach

First of all, it is important to point out that the three prototypes in the present study are prototypes in lay perceptions of personality. They are prototypes in personality descriptions, not prototypes in patterns of objectively recorded behaviour. Although the reported personality patterns surely reflect actual patterns of behaviour to some extent, these descriptions are also influenced by lay theories about personality. Thus the identified prototypes are partly in the eye of the beholders.

Second, the variation of the present four studies in terms of ages, judges, and item samples was not sufficiently systematic to evaluate more subtle differences between the three types with regard to ages, judges, and items. For example, there was no Q-sort study for adults. York and John (1992) reported four instead of three replicable Q-factors, but their sample was restricted to women. Also, our adult studies used self-ratings whereas our

child studies used parental ratings, thus confounding age and type of judge. There may be some differences in the prototypes with regard to self- versus other ratings. For example, neuroticism may be more marked in other-rated undercontrol than in self-rated undercontrol because for other ratings, the less observable trait of neuroticism may be more susceptible to a negative halo effect. Similarly, low agreeableness may be more marked in other-rated undercontrol than in self-rated undercontrol because unfriendliness may be more important for others than for self.

Third, the four studies were all conducted in Germany. We expect that the main findings generalize to other Western cultures, but empirical evidence is lacking in this respect.

Fourth, the clustering method that was applied in the present study and all preceding type studies is based on Ward's method that tends to produce clusters of similar size. Although this property of the procedure is practical for description and prediction, there is no *a priori* reason to assume that personality types do have a similar size. Two observations indicate, however, that this problem is not critical for our results. First, the results of the clustering studies were consistent with the results of the Q-factor studies. Second, the resilient group was clearly larger than the other two groups in all three clustering studies. Thus, the clustering method is sensitive at least to moderately large differences in cluster size.

Directions for future research

Future research on personality types may follow four main directions. First, the influence of culture, ages, judges, personality description, and method of deriving types can be studied systematically by holding four of these factors constant and varying only one. For example, self- and acquaintance NEO-FFI scales can be contrasted with one another, or types based on Q-factors and on Big Five scales that consist of Q-sort item composites can be compared. We expect that small but significant effects will be found for these influence factors, and that the consistency of the prototypes across studies increases even more when these factors are controlled.

Second, the three-type model of personality should be differentiated into a more fine-grainedclassification system by distinguishing subtypes of the three types. Robins, John and Caspi (1998) showed that when the three types identified for late childhood by Q-factor analysis are once more separately Q-factor analysed, replicable factors were found for resilients and undercontrollers that can be interpreted as subtypes of them. Similar subtypes were identified by van Lieshout, Haselager and van Aken (1998) for adolescence through cluster analyses of Big Five ratings in a sample of more than 3000 12–18-year-olds. These first results suggest that it seems promising to search for replicable subtypes of the three types in large samples not only of children but also of adults.

Third, it seems that the long-term stability of the individual membership for the types is not high despite a sufficient short-term retest reliability and despite a high continuity of the prototypes. Consequently, there seem to be meaningful transitions between the types over development that cannot be attributed to their unreliability or to a discontinuity of the classification. Longitudinal analyses are needed that describe these patterns of intraindividual change.

Last but not least, once types, subtypes, and their developmental change have been described in sufficient detail, an important step will be to move from description to explanation. What are the genetic and environmental antecedents of these (sub)types and what are the personal and social consequences of (sub)type differences?

CONCLUSION

This study has shown that patterns of personality description are organized around three major personality prototypes: resilient, overcontrolled, and undercontrolled. The borders between the corresponding personality types are fuzzy but replicable, generalizable across ages and methods, and construct valid. These types provide a more realistic picture of the interindividual structure of intraindividual personality patterns than linear or curvilinear trait models, but only at a coarse level of differentiation. Therefore, the within-type variability of the patterns is still high. However, the three-prototype model can serve as a solid starting point for more differentiated classifications in terms of subtypes.

In our view, personality researchers should take deviations from the traditional linear model seriously and exploit their inherent information about within-personality organization. The prototype approach to personality differences makes this possible. Another advantage is that it facilitates communication with anybody interested in applications of personality description and prediction because differences between types are easier to communicate to clinicians and the general public than correlations with a trait. Therefore, researchers often use the type language to communicate correlational findings by presenting extreme group comparisons, however with arbitrary cut-points along only one or two trait dimensions. The present prototype approach offers distinctions that are less arbitrary because they are backed up by the empirical analysis of many trait dimensions.

ACKNOWLEDGEMENTS

This research was supported by the Max Planck Society and the German Research Foundation. We wish to thank Iain Glen for stylistic corrections.

REFERENCES

Achenbach TM, Edelbrock CS. 1981. Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development* **46**: (1, Serial No. 188).

Allport GW. 1937. *Personality: A psychological Interpretation*. Holt: New York.

Asendorpf JB, Meier GH. 1993. Personality effects on children's speech in everyday life: Sociability-mediated exposure and shyness-mediated reactivity to social situations. *Journal of Personality and Social Psychology* **64**: 1072–1083.

Asendorpf JB, van Aken MAG. 1993. Deutsche Versionen der Selbstkonzeptskalen von Harter [German versions of Harter's self-concept scales]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* **25**: 64–86.

Asendorpf JB, van Aken MAG. 1994. Traits and relationship status: stranger versus peer group inhibition and test intelligence versus peer group competence as early predictors of later self-esteem. *Child Development* **65**: 1786–1798.

Asendorpf JB, van Aken MAG. 1999. Resilient, overcontrolled, and undercontrolled personality prototypes in childhood: replicability, predictive power, and the trait-type issue. *Journal of Personality and Social Psychology* **77**: 815–832.

Asendorpf JB, Wilpers S. 1998. Personality effects on social relationships. *Journal of Personality and Social Psychology* **74**: 1531–1544.

Asendorpf JB, Wilpers S. 2000. Attachment security and available support: closely linked relationship qualities. *Journal of Personal and Social Relationships* **17**: 115–138.

Bartholomew K, Horowitz LM. 1991. Attachment styles among young adults: a test of a four-category model. *Journal of Personality and Social Psychology* **61**: 226–244.
Block J. 1961. *The Q-Sort Method in Personality Assessment and Psychiatric Research*. Thomas: Springfield, IL.
Block J. 1971. *Lives Through Time*. Bancroft: Berkeley.
Block J. 1995. A contrarian view of the five-factor approach to personality description. *Psychological Bulletin* **117**: 187–215.
Block JH, Block J. 1980. The role of ego-control and ego-resiliency in the organization of behavior. In *Minnesota Symposium on Child Psychology,* vol. 13, Collins WA (ed.). Erlbaum: Hillsdale, NJ: 39–101.
Borkenau P, Ostendorf F. 1993. *NEO-Fünf-Faktoren Inventar (NEO-FFI) [NEO-Five-Factor Inventory]*. Göttingen: Verlag für Psychologie.
Breckenridge JN. 1989. Replicating cluster analysis: Method, consistency, and validity. *Multivariate Behavior Research* **24**: 147–161.
Caspi A. 1998. Personality development across the life course. In *Handbook of Child Psychology,* 5th edn, vol. 3, Eisenberg N (ed.), *Social, Emotional, and Personality Development*. Wiley: New York, 311–388.
Caspi A, Moffitt TE, Newman DL, Silva PA. 1996. Behavioral observations at age 3 years predict adult psychiatric disorders. *Archives of General Psychiatry* **53**: 1033–1039.
Caspi A, Silva PA. 1995. Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal evidence from a birth cohort. *Child Development* **66**: 486–498.
Costa PT Jr, McCrae RR. 1989. *The NEO PI/FFI Manual Supplement*. Psychological Assessment: Resources: Odessa, FL.
Digman JM. 1990. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology* **41**: 417–440.
Döring N, Bortz J. 1993. Psychometrische Einsamkeitsforschung: Deutsche Neukonstruktion der UCLA Loneliness Scale [Psychometric loneliness research: a new German version of the UCLA loneliness scale]. *Diagnostica* **39**: 224–239.
Everett JE. 1983. Factor congruence as a criterion for determining the number of factors. *Multivariate Behavioral Research* **18**: 197–218.
Funder DC. 1991. Global traits: a neo-Allportian approach to personality. *Psychological Science* **2**: 31–39.
Gangestad S, Snyder M. 1985. 'To carve nature at its joints': on the existence of discrete classes in personality. *Psychological Review* **92**: 317–349.
Goldberg LR. 1983. The magical number five, plus or minus two: Some considerations on the dimensionality of personality descriptors. Paper presented at a Research Seminar, Gerontology Research Center, NIA/NH, Baltimore, MD.
Goldberg LR. 1992. The development of markers for the Big-Five factor structure. *Psychological Assessment* **4**: 26–42.
Göttert R, Asendorpf J. 1989. Eine deutsche Version des California-Child-Q-Sort, Kurzform [A German short version of the California Child Q-Set]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* **21**: 70–82.
Halverson CF, Kohnstamm GA, Martin RP. 1994. *The Developing Structure of Temperament and Personality from Infancy to Adulthood. Hillsdale NJ, Erlbaum.*
Hart D, Hofmann V, Edelstein W, Keller M. 1997. The relation of childhood personality types to adolescent behavior and development: a longitudinal study of Icelandic children. *Developmental Psychology* **33**: 195–205.
Harter S. 1985. Manual for the self-perception profile for children. Technical Report. University of Denver, CO.
John OP, 1990. The 'Big Five' factor taxonomy: dimensions of personality in the natural language and in questionnaires. In *Handbook of Personality: Theory and Research*, Pervin L (ed.). Guilford: New York, 66–100.
John OP, Caspi A, Robins RW, Moffitt TE, Stouthamer-Loeber M. 1994. The 'Little Five' Exploring the nomological network of the five-factor model of personality in adolescent boys. *Child Development* **65**: 160–178.
John OP, Goldberg LR, Angleitner A. 1984. Better than the alphabet: Taxonomies of personality-descriptive terms in English, Dutch, and German. In *Personality Psychology in Europe:*

Theoretical and Empirical Developments, vol. 1, Bonarius HCJ, van Heck GLM, Smid NG (eds). Swets Zeitlinger: Lisse; pp. 83–100.

Marsh HW, O'Neill, R. 1984. Self description questionnaire III: the construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement* **21**: 153–174.

McCrae RR, Costa PT. 1985. Updating Norman's adequate taxonomy: intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology* **49**: 710–721.

McCrae RR, Costa PT. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology* **52**: 81–90.

McCrae RR, Costa PT. 1990. *Personality in Adulthood.* Guilford: New York.

McCrae RR, Costa PT. 1992. Discriminant validity of the NEO-PI-R facets scales. *Educational and Psychological Measurement* **52**: 229–237.

McCrae RR, Costa PT, Busch CM. 1986. Evaluating comprehensiveness in personality systems: the California Q-set and the five-factor model. *Journal of Personality* **54**: 430–446.

Norman WT. 1963. Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology* **66**: 574–583.

Ostendorf F. 1990. *Sprache und Persönlichkeitsstruktur: Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit [Language and Personality Structure: Validity of the Five-Factor Model of Personality]*. Roderer: Regensburg.

Paunonen SV. 1998. Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology* **74**: 538–556.

Peabody D. 1984. Personality dimensions through trait inferences. *Journal of Personality and Social Psychology* **46**: 384–403.

Peabody D. 1987. Selecting representative trait adjectives. *Journal of Personality and Social Psychology* **52**: 59–71.

Peabody D, Goldberg LR. 1989. Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology* **57**: 552–567.

Pulkkinen L. 1996. Female and male personality styles: a typological and developmental analysis. *Journal of Personality and Social Psychology* **70**: 1288–1306.

Robins RW, John OP, Caspi A. 1998. The typological approach to studying personality. In *Methods and Models for Studying the Individual*, Cairns RB, Bergman LR, Kagan J (eds). Sage: Beverly Hills, CA, 135–160.

Robins RW, John OP, Caspi A, Moffitt TE, Stouthamer-Loeber M, 1996. Resilient, overcontrolled, and undercontrolled boys: three replicable personality types. *Journal of Personality and Social Psychology* **70**: 157–171.

Saucier G, Ostendorf F. 1999. Hierarchical subcomponents of the Big Five personality factors: a cross-language replication. *Journal of Personality and Social Psychology* **76**: 613–627.

van Lieshout CFM, Haselager GJT, Riksen-Walraven JM, van Aken MAG. 1995. Personality development in middle childhood. Paper presented at the Society for Research in Child Development, Indianapolis, IN, 1995.

van Lieshout CFM, Haselager GJT, van Aken MAG. 1998. Generalizability of personality types and subtypes across age, gender, and methods of assessment. Paper presented at the 15th Biennial Meetings of the International Society for the Study of Behavioral Development, Bern, 1998.

Weinert FE, Schneider W. (Eds). 1999. *Individual Development from 3 to 12: Findings from a Longitudinal Study.* Cambridge University Press: Cambridge.

York KL, John OP. 1992. The four faces of Eve: A typological analysis of women's personality at midlife. *Journal of Personality and Social Psychology* **63**: 494–508.

Zetenyi T. (Ed). 1988. *Fuzzy Sets in Psychology.* North-Holland: Amsterdam.

