A Brunswikean Approach to Trait Continuity: Application to Shyness

Jens B. Asendorpf
Max-Planck-Institut für psychologische Forschung
Munich, Germany

ABSTRACT A new method of assessing the continuity of traits independent of their stability is proposed that is based on Brunswik's lens model of person perception. The method is applied to the continuity of shyness between preschool age and adulthood. A sample of 108 children (observed at age 4 and again at age 6) and a sample of 70 university students were videotaped in conversations with adult strangers. Detailed coding of 17 different behaviors revealed a highly similar rank order of their validity coefficients for parental reports of children's shyness, self-reports of students' shyness, and observer judgments of both children's and students' shyness, even after controlling for differences in the behaviors' reliability. These findings suggest that the construct of shyness shows a strong continuity from preschool age through adulthood. The proposed methodology can also be applied to the continuity of emotional states or to the cross-situational generality of traits and states.

INTRODUCTION

In the last two decades, numerous studies have found that interindividual differences in behavior show an increasing temporal stability with increasing age as evidenced by increasing correlations over retest in-

Part of the study reported here was conducted within the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC) funded by the Max Planck Society. I wish to thank the children and parents involved in this study as well as the students who served as subjects or observers, and Jaan Valsiner, Stephen Briggs, and three anonymous reviewers for their comments on earlier drafts of this manuscript. Correspondence concerning this article should be addressed to Jens B. Asendorpf, Max-Planck-Institut für psychologische Forschung, Leopoldstr. 24, D-8000 München 40, Germany.

Journal of Personality 60:1, March 1992. Copyright © 1992 by Duke University Press. CCC 0022-3506/92/$1.50

tervals of equal length. Although the stabilities vary according to the particular trait and the method of observation, in most cases stability coefficients for comparable retest intervals are very low during infancy and increase over time until high stabilities are reached in adulthood (see Moss & Susman, 1980, and Conley, 1984, for reviews). For example, predictions of preschool children's IQ from traditional measures of infants' IQ seldom have overcome the .30 barrier (Kopp & McCall, 1980), whereas Wilson (1983) found a continuous increase of the 1-year stabilities of IQ from .74 (ages 2 to 3) to .90 (ages 8 to 9).

Recently however, studies have found surprisingly high correlations between highly aggregated measures of infants' visual attention and preschool IQ (e.g., .60 between visual attention at age 6 months and the Stanford-Binet IQ at age 3; Bornstein & Sigman, 1986; Rose, Feldman, & Wallace, 1988). This discrepancy between the old and the new stability data for tests that are considered to measure intellectual capacity points to a chronic problem of the interpretation of stability coefficients.

The stability of observed interindividual differences in behavior confounds the *continuity* of the construct underlying the behavior and the *stability* of the individual differences in this construct. A construct is continuous between two points in time if the construct can be operationalized by the same behaviors at both time points. Thus, these behaviors have the same construct validity at both assessments (see MacCorquodale & Meehl, 1948, for the distinction between constructs and empirical indicators, and Cronbach & Meehl, 1955, for the concept of construct validity).

Kagan (1980) called this concept of continuity "homotypic continuity" and distinguished it from "heterotypic continuity" where the empirical indicators of the same construct change over time, and from "discontinuity" where the construct itself applies to one point in time but not to the other. The distinction between heterotypic continuity and discontinuity is a matter of theoretical decision rather than of empirical investigation because in empirical studies constructs are always operationalized in terms of some observables. Thus, for empirical studies of the continuity of traits it suffices to draw a distinction between continuity (homotypic continuity sensu Kagan) and discontinuity (heterotypic continuity or discontinuity sensu Kagan).

If the observed stability is high, it can be safely concluded that the construct is continuous and that the stability of the individual differences in the construct is high. However, if the observed stability is low, this might be due to either a change of the rank order of individuals in the construct or to a discontinuity of the construct.

Low stabilities of interindividual differences in behavior have most often been interpreted as an instability at the construct level, particularly if a high internal consistency and a high short-term stability of the behavioral indicators have been demonstrated. However, this demonstration does not guarantee high construct validity and hence cannot exclude the alternative interpretations that the low observed stability was in fact due to a low construct validity of the assessments at one or both measurement points, or to a discontinuity of the construct. For example, traditional intelligence tests for infancy meet the requirements of internal consistency and short-term retest stability, but it can be claimed that their construct validity for intelligence is poor because visual attention proved to be clearly superior in predicting preschool age IQ.

The aim of the present study is to demonstrate that psychology can profitably move beyond hunting for ever-higher stabilities by enlarging the range of questions that are asked about the stability of personality traits. The continuity of a trait can be studied independently of the stability of the trait. Such an approach provides a firmer base for operationalizations of the same trait at different ages, and for the study of stability.

An early approach to the question of continuity was presented by Emmerich (1964), who defined continuity as the temporal constancy of the factorial structure of a large item pool. If the factor structure remains the same, each factor may be regarded as reflecting a dimension of personality that is continuous over time. Emmerich (1964) interpreted changes in the percentage of variance accounted for by a particular factor as the emergence or decline of this factor. More straightforward are recent approaches of testing the temporal constancy of a particular factor structure by confirmatory factor analysis or structural modeling techniques (see Hertzog & Schaie, 1986, for an application of LISREL to the continuity of the factor structure of intelligence in adulthood).

In these approaches, the continuity of a construct is defined in terms of the relations between this construct and other constructs. The problem is that these relations will change if only one of the constructs is discontinuous. If the relations are very stable over time, the continuity of all constructs can be taken for granted; but if they are unstable, it can be difficult or even impossible to decide which constructs' discontinuity caused the instability of the nomological network.

The approach of the present study is radically different from this traditional way of conceptualizing continuity. Constructs are not reconstructed from intercorrelations of empirical indicators. Instead, a

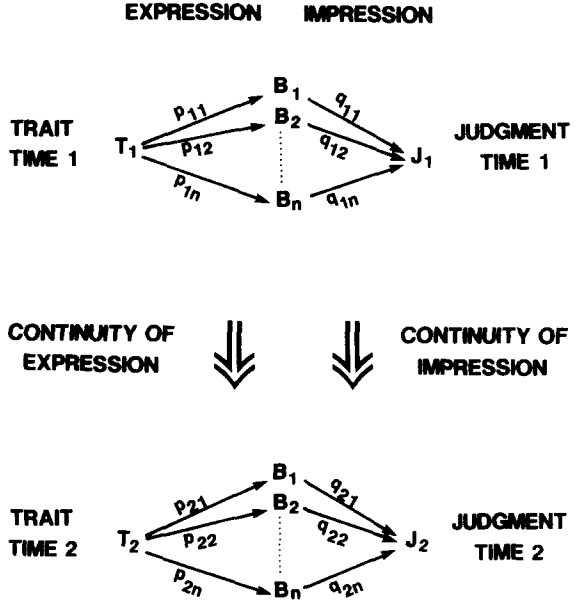

Figure 1
A Lens Model of the Continuity of the Expression and Impression of Traits in Terms of a Sample of Behaviors $B_1, \ldots, B_n$

construct is defined and measured separately from behavior by some criterion measure. This criterion measure is correlated with frequency or intensity measures of many behaviors. The behaviors are not exclusively selected for high correlations with the criterion. Instead, behaviors are sampled "representatively," that is, in a way that represents the whole continuum from very valid to completely invalid behaviors. The temporal stability of the rank order of the validity coefficients of these behaviors reflects the continuity of the construct.

This approach to the continuity of personality traits follows the perspective of functional probabilism (Brunswik, 1955, 1956). Although Brunswik himself did not apply this approach to person perception, Heider (1958), Scherer (1978), and Asendorpf (1990a) did (see also Petrinovich, 1979, for a general discussion of functional probabilism). Figure 1 presents a general model of the continuity of personality traits that draws heavily upon Brunswik's lens model of perception.

It is assumed that the expression of a trait in behavior is reflected by relations of different strength with the intensity or frequency of behaviors that can be observed in trait-relevant situations. The strength

of the relations between the behaviors and the trait can be measured by validity scores (e.g., correlations). Thus, the expression of a trait in behavior is represented not by a few highly valid indicators but rather by the validity structure of a whole sample of behaviors. The behaviors serve as "distal cues" (Brunswik, 1956) for the behavioral analysis of the trait which is based on objective coding of behavior.

Also, the behaviors serve as distal cues for the impression of naive observers who infer a trait judgment from some of these behaviors. Thus, the behaviors may also be ordered in terms of their saliency scores for this trait judgment. The terms "expression" and "impression" are used here because they express the symmetry of these two concepts and because they are frequently used in lay psychology in similar contexts. Brunswik (1955, 1956) preferred the more artificial terms "externalization" and "inference" in his work, probably because at his time "expression" and "impression" were used in a quite sloppy sense in the German "*Ausdruckspsychologie*" (psychology of expression; see Asendorpf & Wallbott, 1982, for a review). Today there is no need of avoiding these terms anymore (see, e.g., Asendorpf, 1990a, and Zivin, 1985, for modern concepts of expression and impression).

In Figure 1, the validity scores $p_{t1}, \ldots, p_{tn}$ and the saliency scores $q_{t1}, \ldots, q_{tn}$ of a sample of n behaviors are contrasted for the same trait at two different points, $t = 1, 2$, in time. The stability of the validity scores between these two points in time is regarded as a measure of the continuity of the expression of the trait. Accordingly, the temporal stability of the saliency scores is considered to be a measure of the continuity of the impression of the trait. The two time points refer to different ages of individuals whose trait expression or impression is evaluated for continuity.

This very general model of the continuity of personality traits must be specified in three respects in order to be applicable in empirical studies. First, it must be decided how traits are measured independently of their expression in behavior. The answer depends, of course, on the type of trait under study. Any characteristic of a person may be used as a trait criterion, such as self-ratings or ratings of knowledgeable informants, but more "objective" characteristics of individuals, such as genetic dispositions or physiological response tendencies, also may be used. It is only necessary to ensure that the criterion measure is assessed independently of the behaviors included in the model. If ratings serve as the trait criterion, the distinction between expression and impression is reduced to the fact that these ratings reflect global judgments that are

obtained independently of the behavioral assessment and hence depend only indirectly on these behaviors, whereas the impressions depend on them directly.

Second, the behaviors must be carefully sampled. Psychologists are well aware that subject sampling is crucial for the results obtained, but they often pay much less attention to the fact that the same is often true for the sampling of situations and behaviors (Brunswik, 1956, was one of the first who acknowledged this point). It is easy to misuse the present model if the behaviors are selected without care or in a way that artificially increases or decreases continuity.

Continuity as defined in the present model depends strongly on the variance of the validities of the behaviors for the trait, or on the variance of the saliencies of the behaviors for the trait judgment. These variances could be artificially increased by adding to each highly valid expression the inverted, but otherwise identical, variable. For example, if the percentage of speech correlates .50 with the trait criterion, one might also include the percentage of silence into the list of behaviors, although this behavior correlates -1 with the percentage of speech; this redundant addition will artificially increase the variance of the validities and hence also the continuity of the trait expression. Also, the variance of validities could be artificially decreased by neglecting highly valid behaviors, or by adding many completely invalid behaviors. Just as the stability coefficients for individual differences vary with the selection of individuals, the continuity coefficients of the present model vary with the selection of both individuals and behaviors.

Third, the validities and saliencies of behaviors partly depend upon the degree to which the behaviors reliably reflect interindividual differences. The reliability of a behavioral indicator of a trait is an upper limit of the validity of the indicator for this trait. If the behaviors vary greatly in their reliability, the temporal stability of this variation affects the continuity of the expression and impression of the trait. If the differences in reliability are highly stable over time, trait continuity will be inflated; if they show negative correlations over time, continuity will be underestimated. Thus, the reliability differences among the behaviors and their possible influence on trait continuity must be evaluated.

To summarize, two major aspects of the continuity of a trait are distinguished here: the temporal stability of the validity coefficients of behavioral indicators of this trait, and the temporal stability of the saliency of these indicators for the impression formation about the trait. Note that this approach to trait continuity can also be applied to the con-

tinuity of emotional or motivational states, or to the cross-situational generality of traits and states.

In the present study, the model of continuity developed so far was applied to the trait of shyness. Shyness seemed to be a good candidate for applying the model because (*a*) shyness appears to be a major dimension of personality both among young children and among adults of all ages (see Jones, Cheek, & Briggs, 1986), (*b*) shyness is a personality trait that is deeply rooted in lay psychology and hence easy to assess in terms of self- and other-ratings (see Jones, Briggs, & Smith, 1986, for the strong convergence of different shyness scales), and (*c*) shyness is an easily observable trait and hence high in terms of self-other agreement (see Funder & Colvin, 1988, for empirical evidence for the influence of trait observability on self-other agreement). A high validity of the judgment of others is particularly important if, as in the present study, traits are assessed in early childhood when valid self-ratings of personality cannot be obtained.

An ideal empirical test of the continuity of shyness between childhood and adulthood requires comparison of the reactions accompanying shyness in preschool children with their reactions as adults by applying exactly the same method of behavioral analysis to their behavior in exactly comparable situations. The present study does not fully meet this ideal because it consists of reanalyses of existing studies (Asendorpf, 1987, 1990b) that refer to similar observational settings and even identical methods of behavioral analysis but that originally were not designed to test the continuity of shyness.

Ideally, the same subjects should be studied. However, this would require a longitudinal study in which the same methods of behavioral analysis would have been applied to a time span of 20 years. Alternatively, independent samples can be used for different ages. If a high continuity is found, the variation of the subject sample even increases the strength of the results. The present study is based on a mixed design. The continuity of shyness was evaluated by comparing a longitudinal sample of children, who were assessed both at age 4 and at age 6, and a different sample of university students (average age, 23 years).

As was pointed out above, the sampling of behaviors is crucial for applications of the proposed model of continuity. In the present study, subjects' speech, gaze, and body movement were systematically observed because a literature review of the behavioral expression of shyness suggested that in each of these three nonverbal domains some variables would be related to shyness (Asendorpf, 1990a). However,

not only behaviors that were expected to be valid for shyness were analyzed within these nonverbal domains. Instead, each domain was systematically sampled by categorizing all possible behaviors in the domain into disjunct and exhaustive classes. For example, rather than selecting subjects' percentage of silence as the most promising indicator from the domain of speech, speech behavior was systematically sampled by distinguishing four states of speech (cross-classification of subjects' and partners' speech vs. no speech). Furthermore, one additional variable that seemed to be a particularly powerful predictor of shyness was added to this list: the latency until the first spontaneous (unsolicited) utterance to a partner (see Daly, 1978; Kagan, Reznick, & Snidman, 1987).

These behavioral indicators of shyness were correlated with self-ratings of shyness (for students) and parental ratings of shyness (for preschool children) that were obtained independently of the behavioral assessment. It was necessary to apply different criterion measures of shyness because self-ratings of preschool children lack validity. Finally, the continuity of the impression formation of naive adult observers about the shyness of children and adults was analyzed by correlating their judgments of shyness with the behaviors selected for the study.

METHOD

Subjects

Two samples are compared in this study. A sample of 108 children born between August 1980 and July 1981 was observed at the age of 3;5–4;5 years (M = 3;11 years), and 2 years (±2 months) later at the age of 5;5–6;5 years (M = 5;11 years). This sample was part of the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC), which follows a rather unbiased sample of children of the Munich area in terms of IQ and social class (see Asendorpf, 1990b; Weinert & Schneider, 1989). Of the 108 children (58 boys, 50 girls), 88 were observed at the first assessment, 90 at the second assessment, and 70 at both assessments.

The adult sample consisted of 70 subjects age 19 to 32 years (M = 23 years; 35 males, 35 females) who were selected from a sample of 307 university students from the upper quartile and the lower tercile of a shyness self-rating scale (see below). Thus, this sample was biased toward high and low shyness.

Observational Settings

Both children and adults were observed during a conversation with an adult stranger who was trained to react in a standardized way to the subjects. All conversations took place at a table in the same room, were videotaped through a one-way mirror, and involved the same female stranger for all 4-year-olds, another female stranger for all 6-year-olds, the same female confederate for all female students, and the same male confederate for all male students; these confederates pretended to be subjects in the study.

For the children, the conversation started with a pre-interaction period during which the stranger responded to initiation attempts of the child but did not actively approach the child. This procedure was aimed at inducing stranger anxiety. If the child did not initiate a conversation with the stranger within 3 minutes, the stranger tried her best to involve the child in a conversation. In any case, there was a pre-interaction period of 0 to 3 minutes until either the child or the stranger initiated the conversation, and an interaction period of 2 minutes following this first initiation (see Asendorpf, 1990b, for more details).

For the students, the conversation started with a waiting period of 3 minutes during which both partners ostensibly waited for the "real experiment" without an experimenter being present. Following this situation, which was aimed at inducing stranger anxiety, the experimenter set up a video camera in the room and instructed both partners to get to know each other in order to evaluate their partner's personality later on in a questionnaire. Subsequently, the experimenter left the room for 3 minutes. This procedure was aimed at inducing social-evaluative anxiety (see Asendorpf, 1987, for more details). All situations of all three assessments were videotaped.

Measures

Criterion measures of shyness. The parent who accompanied the child in the Institute (nearly always the mother) answered a questionnaire which contained 4 items to be rated on a 7-point scale ("never" to "always") that referred to shyness toward adult strangers (e.g., "My child is shy toward unknown adults"). These 4 items were randomly distributed among 44 other items of the same response format. The internal consistency of this shyness scale was very high (Cronbach's $\alpha = .95$ for the 4-year-olds and .93 for the 6-year-olds); the 2-year stability was also high ($r = .70$, $n = 70$).

About 2 months before the observation, the students answered a questionnaire which contained 4 items to be rated on the same 7-point scale that was also used for the parental ratings. These items referred to shyness toward adults (e.g., "I feel shy in the presence of others"; see Asendorpf, 1987, for all items). These 4 items were randomly distributed among 44 other items of the

same response format. The internal consistency of this shyness scale was high (Cronbach's $\alpha = .79$). Thus, the parental scale and the student scale were quite comparable.

Observer ratings of shyness. All videotapes of an assessment were presented to the same group of observers. Different students served as observers for different assessments (four observers for the 4-year-olds, two for the 6-year-olds, and three for the students). The observers rated the subjects' shyness after each minute of recording on a 7-point scale ("not at all" to "very much") labeled "shy-inhibited." For children, the observers watched the 2 minutes of interaction following the first initiation; for students, they watched all 6 minutes of interaction. All observer ratings were averaged, yielding one shyness score per subject and observer. These scores showed a high inter-observer agreement (Cronbach's $\alpha = .97$ for the 4-year-olds, .89 for the 6-year-olds, and .84 for the students). The mean of the shyness scores of all observers served as the observer rating of shyness.

Behavioral coding. The recordings presented to the observers were coded by different students for the onset and offset times of the following variables: speaking by the subject, speaking by the subject's partner, the subject's gazing at the partner, self-adaptors of the subject (touching one's face or neck with one's hand), and illustrators of the subject (hand movements illustrating speech) (see Ekman & Friesen [1972], for a discussion of hand movements). Because the measures of gazing and hand movements were of limited success for predicting shyness among the 4-year-olds, gazing and hand movements of the 6-year-olds were not coded; therefore the behavior sample for the 6-year-olds is somewhat restricted. Because illustrators were very rare among the 4-year-olds, they are not considered in the present analysis.

All codings were done on a microcomputer that was synchronized with the video recorder by an interface. Thus, onset and offset times were simply coded by pressing a button. Speech was monitored continuously in the normal play mode of the video recorder. Coding times were corrected for the mean reaction time of the coder (0.4 seconds) that had been determined by a comparison of one full coding protocol, with the coding times determined exactly by slow-motion analysis. Gazing and self-adaptors were always coded by slow-motion analysis. Thus, all coding times referred to the same time base, and codes of different coding systems could be combined.

The onset and offset times were transformed into interval codes with an interval length of 0.3 seconds. From these time sequences, the frequencies and durations of each code, as well as of combinations of different codes, were computed, and the frequencies were transformed into mean lengths (mean length = duration/frequency) because for conversational behaviors, mean length seems to be more meaningful than frequency (see Duncan & Fiske,

1977). The subjects' and their partners' on-off patterns of speech as well as subjects' speech and gazing were systematically combined, yielding the following codes: speaking alone, listening, silence, double-talk, gazing during speaking, and gazing during silence. Thus, the measures of gaze were not intrinsically related to the measures of speech.

In addition to these "content-free" codings of nonverbal behavior, an additional behavior was coded that depended upon the verbal content of subjects' speech. For each subject, the latency to the subject's first spontaneous (unsolicited) utterance was determined because this behavioral measure seems to be a particularly powerful indicator of shyness (see Introduction). For children, these latencies were approximately normally distributed except for a strong ceiling effect (the maximum latency of 300 seconds was obtained by 41% of the 4-year-olds and by 12% of the 6-year-olds). Therefore, correlations with these latencies were corrected for this ceiling effect (see Alliger, Hanges, & Alexander, 1988). For students, the latencies showed no ceiling effect.

RESULTS

Intercoder Agreement

In addition to the main coders who coded all tapes of an assessment either for speech, for gaze, or for self-adaptors, a second student independently coded 9% to 10% of the tapes of each assessment. The main coders did not know which of their tapes would be selected for this reliability check. Intercoder agreement was evaluated based on the interval codings generated by the two coders for the same material. Table 1 presents the intercoder agreement for the three assessments and the three coding systems.

Table 1 indicates that the codings were reliable; in each case, they exceeded the conventional criterium of $\kappa = .60$. The relatively low κ of the coding of the very young children's speaking, as compared to the other κs or to the percentage of agreement, appeared to be due to the short mean length of speech of these children (0.6 seconds on average) that made the coding of speech more susceptible to reaction time errors.

The reliability of the latency to children's first spontaneous utterance was determined by having this latency coded by a second coder for 30 subjects of each assessment, and by correlating the two codings across subjects. After correcting children's data for the ceiling effect (see Method section), these correlations were satisfactory (.93 for the 4-year-olds, .94 for the 6-year-olds, and .84 for the students). The cor-

Table 1
Intercoder Agreement

	Children				Students	
	4 years		6 years		19 to 32 years	
Behavioral coding	%	κ	%	κ	%	κ
Speaking by subject	94	.63	95	.69	93	.81
Speaking by partner	88	.69	90	.71	92	.80
Gazing by subject	95	.71	—	—	96	.92
Self-adaptors by subject	99	.93	—	—	99	.98

Note. % refers to number in agreement/(number in agreement + number in disagreement); κ is Cohen's Kappa.

relation for the students was lower because the variance of the latencies was much smaller (maximum 90 seconds).

Short-Term Stability of Behaviors

The short-term stability of each behavior was determined by computing Spearman rank order correlations between the first half of each situation with the second half, and by applying the Spearman-Brown formula to these correlations (split-half reliability). Spearman correlations were used because some of the measures had highly skewed distributions which inflate Pearson correlations and Cronbach's α. Obviously, the latency measure could not be evaluated for short-term stability.

Table 2 presents the short-term stability of the behaviors in terms of their split-half reliability (except for the latency scores, for which no reliability could be determined). Table 2 indicates that the reliabilities varied considerably both within and between assessments. Mean reliabilities were .53 (4-year-olds), .45 (6-year-olds), and .64 (students). These reliabilities may underestimate true reliability for children because they are based on two situations of only 1 minute of duration. The longer observation periods, as well as the extreme group selection of subjects, might explain the higher reliability for the students.

For the present type of analysis, the mean level of reliability is less relevant than differences among the reliabilities of the measures and the stability of these differences between different age groups. This stability was evaluated by computing Spearman correlations between the three columns of Table 2. The correlation between the reliabilities of children was high (.74, $p < .04$), whereas the two correla-

Table 2
Reliabilities of Individual Differences in Children's and Adults' Behavior

	Split-half reliability[a]		
	Children		Students
Behavioral measures	4 years[b]	6 years[c]	19 to 32 years[d]
Percentage speaking alone	.84	.66	.52
Mean length speaking alone	.55	.34	.73
Percentage listening	.25	.40	.25
Mean length listening	.33	.35	.43
Percentage silence	.52	.46	.80
Mean length silence	.63	.60	.71
Percentage double-talk[e]	.70	.48	.84
Mean length double-talk	.28	.30	.44
Percentage gazing during speaking	.31	—	.84
Mean length gazing during speaking	.70	—	.68
Percentage gazing during silence	.73	—	.80
Mean length gazing during silence	.27	—	.59
Mean length gaze aversion during speaking	.67	—	.72
Mean length gaze aversion during silence	.78	—	.72
Percentage self-adaptors	.56	—	.63
Mean length self-adaptors	.29	—	.54

a. Spearman-Brown-corrected Spearman correlations between first half and second half of conversation.
b. $N = 88$.
c. $N = 91$.
d. $N = 70$.
e. Percentage of observation time when both partners were speaking at the same time.

tions between students' and children's reliabilities were positive but not significantly different from zero (.41 for the 4-year-olds; .31 for the 6-year-olds). Thus, reliability differences between the behavioral measures were stable within childhood but unstable between childhood and adulthood.

Continuity of the Expression of Shyness

Table 3 presents the correlations between the behavioral measures and the criterion measures of shyness for the three age groups. Table 3

Table 3
Behavioral Correlates of Shyness in Children and Adults

	Children			Students	
	4 years[a]		6 years[b]	23 years[c]	
Behavioral measures	r[d]	Rank[e]	r[d]	r[d]	Rank[e]
Latency to first spontaneous utterance	.64***	1	.50***	.46***	1
Percentage speaking alone	−.55***	17	−.50***	−.40***	17
Mean length speaking alone	−.14	10	−.45***	−.11	10
Percentage listening	−.21	14	−.26*	−.09	9
Mean length listening	.17	5	−.21*	.07	5
Percentage silence	.55***	3	.45***	.41***	2
Mean length silence	.59***	2	.41***	.38**	3
Percentage double-talk[f]	−.40***	16	−.37***	−.33**	15
Mean length double-talk	−.15	13	−.16	−.14	14
Percentage gazing during speaking	−.09	8	—	−.13	13
Mean length gazing during speaking	−.01	7	—	−.07	8
Percentage gazing during silence	−.35**	15	—	−.35**	16
Mean length gazing during silence	−.11	9	—	.04	7
Mean length gaze aversion during speaking	−.15	12	—	.05	6
Mean length gaze aversion during silence	.39***	4	—	.31*	4
Percentage self-adaptors	−.14	11	—	−.11	12
Mean length self-adaptors	.16	6	—	−.11	11

a. $N = 88$.
b. $N = 90$.
c. $N = 70$.
d. Spearman correlations with parental scale (children) or with shyness scale (students) except for children's latencies, for which Pearson correlations are reported that were corrected for the ceiling effect in the latencies; see Method section.
e. Rank of correlation among all behaviors.
f. Percentage of observation time when both partners were speaking at the same time.
*$p < .05$
**$p < .01$
***$p < .001$.

shows a pattern of correlations that was highly consistent across the three age groups. The most powerful predictor of shyness for all age groups was the latency to the first spontaneous utterance. Second came the mean length or the percentage of silence in the dyadic conversations. Similarly predictive, but inversely related to shyness, was the percentage of subjects' speech. Less predictive for shyness were the percentage and the mean length of gaze aversion during silence (but not during speech) and, inversely related to shyness, the percentage of double-talk and the percentage of gazing during silence. Note that, for the question of continuity, it is irrelevant which of these differences within age groups were significant. What is important here is the similarity of the rank order of behaviors across age groups.

If the 17 variables are compared across age in terms of the significance of their validity coefficients for shyness, the significance pattern for 4-year-olds and for university students was identical. However, this identity does not necessarily speak to the issue of continuity. In fact, the significance of a validity coefficient or the significance of cross-age differences in validity coefficients is irrelevant for the question of continuity. Continuity refers to the cross-age similarity of the within-age rank order of validity coefficients, and this similarity is independent of age differences in the mean level of (positive or negative) validity coefficients.

Continuity of the Impression of Shyness

Table 4 presents the correlations between the behavioral measures and the observer ratings of shyness for the three age groups. Again, the correlations were often consistent across age groups, although there were also some inconsistencies between 4-year-olds and students for measures of gaze and self-adaptors. Obviously, some of the correlations were higher than the comparable ones of Table 3 because the observers based their judgment on the very behavior that was also used for coding. Less obvious is the fact that only the shyness-related speech measures showed consistently higher correlations with the impression of shyness; thus, observers apparently relied particularly on speech cues in their judgment.

The rank order of the behavioral measures in terms of their predictive power for the observers' impression of shyness was similar to the one found for the expression of shyness, with two exceptions. First, the latency measure was highly correlated with the observers' impression

Table 4
Behavioral Correlates of Adults' Impressions of Children's and Adults' Shyness

	Children			Students	
	4 years[a]		6 years[b]	23 years[c]	
Behavioral measures	r[d]	Rank[e]	r[d]	r[d]	Rank[e]
Latency to first spontaneous utterance	.81***	1	.77***	.44***	3
Percentage speaking alone	−.81***	17	−.81***	−.66***	17
Mean length speaking alone	−.39**	15	−.58***	−.31**	15
Percentage listening	−.03	8	−.32**	.00	9
Mean length listening	.35***	6	−.29**	.08	5
Percentage silence	.59***	3	.64***	.63***	1
Mean length silence	.71***	2	.61***	.58***	2
Percentage double-talk[f]	−.69***	16	−.62***	−.32**	16
Mean length double-talk	−.03	9	−.21	−.09	11
Percentage gazing during speaking	−.24	13	—	.04	7
Mean length gazing during speaking	−.19	12	—	−.12	12
Percentage gazing during silence	−.14	11	—	−.23	14
Mean length gazing during silence	.12	7	—	.13	4
Mean length gaze aversion during speaking	−.38**	14	—	.00	8
Mean length gaze aversion during silence	.49***	4	—	.05	6
Percentage self-adaptors	−.13	10	—	−.05	10
Mean length self-adaptors	.44*	5	—	−.16	13

a. $N = 88$.
b. $N = 90$.
c. $N = 70$.
d. Spearman correlations with observer ratings of shyness except for children's latencies, for which Pearson correlations are reported that were corrected for the ceiling effect in the latencies; see Method section.
e. Rank of correlation among all behaviors.
f. Percentage of observation time when both partners were speaking at the same time.
*$p < .05$
**$p < .01$
***$p < .001$.

for children but not for adults, whereas it was the most powerful predictor for the shyness criterion of both children and adults. The comparably low correlation with the observer ratings for students is not surprising because children's latency to their first unsolicited utterance was highly variable, covering the whole observation period of 2 minutes, whereas students' latency varied between 0 seconds and 90 seconds, with a mean of 15 seconds, and hence influenced the observers' judgment only during the first seconds of an observation period of 6 minutes.

Second, the mean length of subjects' speech was related to the impression, but not to the expression, of shyness for students and 4-year-olds. This difference may reflect a perceptual bias of the observers who apparently regard frequent short utterances as indicating shyness, although students' self-ratings do not support this bias.

Consistency of the Validity Coefficients of Behaviors across Age and Different Criteria of Shyness

In a meta-analysis of the data presented in Tables 3 and 4, the validity coefficients of the 17 behavioral measures were correlated across the six columns of these tables (Spearman correlations; see the above-diagonal coefficients of Table 5). The correlations between the columns of Table 3 measure the continuity of the behavioral expression of shyness between the three age groups, whereas the correlations between the columns of Table 4 assess the continuity of the impression formation about shyness.

Because the reliabilities of the behaviors showed some degree of cross-age stability, differences among these reliabilities might influence the stability of the validity coefficients. Therefore, in an additional analysis, each validity coefficient was corrected for attenuation by dividing it by the square root of the behavior's split-half reliability. These corrected validities were also correlated across age groups (see the below-diagonal coefficients of Table 5).

The triangles marked in Table 5 present an evaluation of the continuity of the expression of shyness (upper triangle) and of the impression of shyness (lower triangle). For the expression of shyness, mean continuity was .85 (after correction for attenuation, .74); for the impression of shyness, mean continuity was equally high (.88; corrected, .86). Thus, both the expression and the impression of shyness showed a high continuity between 4 years of age and adulthood.

Table 5
Consistency of the Validity Coefficients of 17 Behaviors across Age Groups and Different Criteria for Shyness

Validity		1	2	3	4	5	6
Parental scale for 4-year-olds	1		.85	.84	.83	.90	.78
Parental scale for 6-year-olds	2	.71		.87	.94	.98	.88
Self-ratings of students	3	.84	.67		.76	.90	.86
Observer ratings of 4-year-olds	4	.70	.88	.59		.95	.79
Observer ratings of 6-year-olds	5	.81	.98	.76	.93		.90
Observer ratings of students	6	.66	.90	.80	.73	.93	

Note. Reported are Spearman correlations between the validity coefficients of the columns of Tables 3 and 4 (above diagonal, correlations between raw validities; below diagonal, correlations between validities corrected for attenuation).

The two diagonals of the upper-right and the lower-left quadrant of Table 5 contain the correlations between the validity coefficients for the expression and the impression of shyness. These correlations were comparably high (mean consistency, .89; after correction for attenuation, .83). This consistency indicates that the observers based their ratings on behavioral criteria for shyness that were also used by parents and students for their other- and self-judgments.

Detailed Analysis of Pause Behavior

The behavioral measures of shyness discussed so far are rather global ones. Because subjects' pause behavior appeared to be crucial for shyness according to Table 3, it was analyzed in more detail. All pauses of a subject were classified according to their length. In order to reach comparability across age groups, as well as to avoid differences in predictability due to disproportionate class sizes, a classification procedure was needed that led to class sizes of a comparable frequency both across different pause classes within a subject and across different age groups for a given pause class. A classification procedure grouping the subjects' pauses into 4 intervals with the same logarithmic length (pause length below 0.6 seconds, 0.6 seconds to 1.2 seconds, 1.2 seconds to 2.4 seconds, and above 2.4 seconds) appeared to be the best solution to this problem. Table 6 presents the means and standard deviations of the ipsatized frequencies of these four pause types, as well as the correlations between these ipsatized frequencies and the criterion measure of shyness for each age group.

Table 6
Ipsatized Frequency of Pause Types as Correlates of Shyness in Children and Adults

	Means (standard deviations)[a]			Correlation with shyness[b]		
	Children		Students	Children		Students
Pause length (L)	4 years[c]	6 years[d]	19 to 32 years[e]	4 years[c]	6 years[d]	19 to 32 years[e]
$L \leq 0.6$ seconds	29.0 (9.8)	32.3 (11.9)	52.1 (8.5)	−.37***	−.44***	−.31*
0.6 seconds $< L \leq 1.2$ seconds	24.0 (7.9)	21.6 (7.4)	24.2 (4.8)	−.38***	−.20	.00
1.2 seconds $< L \leq 2.4$ seconds	22.3 (8.2)	24.2 (7.4)	13.2 (4.8)	.03	.18	.00
2.4 seconds $< L$	24.7 (11.2)	21.9 (13.3)	10.5 (5.6)	.57***	.41***	.46***

a. Means (standard deviations) refer to the percentage of each pause type among all pauses of a subject (ipsatized frequency).
b. Pearson correlation between ipsatized frequency of pause type and parental shyness scale (children) or self-rated shyness (students).
c. $N = 88$.
d. $N = 91$.
e. $N = 70$.
*$p < .05$
***$p < .001$.

A similar pattern of correlations was found for all three age groups. The ipsatized frequency of short pauses (up to 0.6 seconds) that typically appear within utterances was inversely related to shyness, the frequency of pauses of medium length (1.2 seconds to 2.4 seconds) was not related to shyness, and the occurrence of long pauses (above 2.4 seconds) that typically appear between different utterances was positively related to shyness. Thus, shyness was characterized in all age groups both by a lack of short pauses within utterances and by a preponderance of long pauses between utterances among the pauses of a subject.

Because short pauses within utterances and long pauses between utterances have a different psychological meaning, this result is not simply a consequence of the higher overall mean length of the pauses of the shy subjects. Instead, it suggests that shy people let long pauses develop and speak less fluently than nonshy individuals.

DISCUSSION

This study investigated the continuity of the expression and impression of shyness between the age of 4 and 6 years, and adulthood. Subjects were videotaped in conversations with strangers, and 17 measures of social-interactional behavior were assessed for each subject. These behaviors were correlated with criterion measures of shyness that were obtained independently of the observational setting (global self- or parental ratings); these correlations describe how interindividual differences in shyness are expressed in behavior among children and adults. Furthermore, the 17 behaviors were correlated with global judgments of shyness obtained from naive observers who watched the subjects' videotaped behavior; these correlations describe how naive observers form their impression about interindividual differences in the shyness of children and adults.

The rank order of the validity coefficients of the 17 behaviors for the criterion measures of shyness, as well as for the observer judgments, was highly similar across all three age groups even if the validity coefficients were controlled for the unreliability of the behaviors in terms of their short-term stability. This similarity indicates that the lay concept of shyness as used in self- and other-judgments refers to similar behavioral criteria at preschool age, kindergarten age, and early adulthood.

This continuity of the lay concept of shyness suggests that it is meaningful to define a psychological construct of shyness in terms of the

same behavioral criteria at these quite different ages. Up to now, researchers had to use common sense when they decided whether to operationalize a particular construct at different ages by the same or by different behaviors. The methodology proposed in the present study allows researchers to base this decision on empirical information.

That shyness shows such a strong continuity may be one of the reasons why shyness, particularly when operationalized as shyness toward strangers, shows considerable stability over time beginning early in childhood. In a meta-analysis of 18 longitudinal studies of shyness during childhood and adolescence, Asendorpf (1989) found an average 5-year stability that increased from around .30 when the first assessment was in infancy to .50 when it was at age 16. We have recently reassessed the children of the present sample at age 8 in a conversation with an adult stranger. Children's shyness at age 4 as measured by the aggregate of the parental scale and the latency measure (see Method section) correlated .75 with the same measure at age 6 (see also Asendorpf, 1990b), and .62 with the same measure at age 8. This 4-year stability of .62 between ages 4 and 8 that was based on the same measure of shyness would be difficult to achieve if the construct of shyness were less continuous.

Traits other than shyness may show a lower degree of continuity. For example, it is unlikely that attachment to a caregiver at the age of 18 months, 4 years, 13 years, and 85 years shows a strong continuity of behavioral expression. Attachment might be a case where it is useful to operationalize the same construct at different ages by different behaviors. The methodology proposed here can help to provide an empirical base for such a decision.

If a lack of continuity is found, a more detailed analysis in terms of particular behaviors can be performed. If the lack of continuity is due to one or two behaviors which change very much in their validity, and if continuity is much higher after disregarding these behaviors, this finding would suggest that these few critical behaviors should not be considered age-independent indicators of the construct under study. They might be highly valid indicators for a particular age, but because of their strong change in validity over age,they are not appropriate for longitudinal analyses of the construct.

The proposed methodology of studying continuity is limited because it refers to behaviors that can be studied in all age groups under consideration. Often, however, very powerful age-specific empirical indicators may exist that cannot be assessed at a different age. For example,

Huesmann, Eron, Lefkowitz, and Walder (1984) studied the stability of interindividual differences in aggressiveness between ages 8 and 30. At age 30, one of the best indicators of aggressiveness was the number of convictions for driving while intoxicated. For 322 males, this index correlated .29, $p < .001$, with aggressiveness ratings by classmates at age 8. Because 8-year-olds rarely consume alcohol or drive cars, it is impossible to investigate the appropriateness of this index for measuring aggressiveness at age 8.

This problem can be solved by a two-step procedure. First, one can search for behavioral domains that appear to be related to the construct under study and that can be measured in all age groups under consideration. If such behavioral domains exist, various behavioral indices can be systematically construed for these domains, and the continuity of the construct can be evaluated using the methodology proposed here. Second, one can search for additional, age-specific indicators and compare them with the behaviors of the continuity analysis in terms of the cross-age predictability of interindividual differences.

If the predictability between age-specific indicators is not higher than the stability of some of the behaviors of the continuity analysis, age-specific indicators should not be used in cross-age analyses because they contribute age-specific error variance, even if their concurrent validity is higher. More interesting is the case that age-specific indicators have more predictive power than those used in the continuity analysis. This finding would indicate that the sample of behaviors for the continuity analysis might be improved by including other behaviors with a higher cross-age stability. The practical problem of this second step is that it needs a longitudinal study, whereas cross-sectional analyses are sufficient for the continuity analysis.

The methodology proposed here can be applied not only to the continuity of traits but also to the continuity of the expression and impression of affective states. For example, children's and adults' behavioral reaction to a frightening or a happy event can be studied cross-sectionally with the same method of behavioral analysis, and then be evaluated for the continuity of the expression and impression of fear or happiness as they are reflected in *intra*-individual changes between a baseline situation and the emotional situation.

Furthermore, the methodology can also be applied to the cross-situational generality of traits or states. For example, the cross-situational generality of the behavioral expression of attachment to parents versus peers could be investigated among a particular age group of children.

If a low cross-partner generality is demonstrated for an appropriate sample of behaviors, the cross-partner generality of the construct of attachment might be questioned, and findings of a low cross-partner consistency of the interindividual differences in particular measures of attachment would be interpreted as reflecting a low generality of the construct rather than a low consistency of the interindividual differences in the same construct.

Future studies should evaluate the continuity and cross-situational generality of various traits in order to explore whether the proposed methodology can be applied to traits other than shyness, and whether the continuity and cross-situational generality differs between different traits. Because only cross-sectional studies are needed, already existing data sets for a particular age group or situation can be expanded by studying other age groups or other situations with the same methods that were applied in the available data.

CONCLUSION

The present study has demonstrated that the continuity of traits can be empirically investigated independently of the stability of these traits as well as independently of the continuity of other traits. Within a Brunswikean lens model of person perception, continuity is defined by the cross-age stability of the validity structure of systematically sampled behaviors. If this validity structure refers to naive trait judgments (as in the present study), the continuity analysis evaluates the continuity of judgment-behavior relationships. The results of such an analysis can be profitably used within a bootstrapping procedure that aims at refining psychological constructs by the continuous interplay between lay judgments and detailed behavioral analyses.

REFERENCES

Alliger, G. M., Hanges, P. J., & Alexander, R. A. (1988). A method for correcting parameter estimates in samples subject to a ceiling. *Psychological Bulletin*, **103**, 424–430.

Asendorpf, J. B. (1987). Videotape reconstruction of emotions and cognitions related to shyness. *Journal of Personality and Social Psychology*, **53**, 542–549.

Asendorpf, J. B. (1989). *Soziale Gehemmtheit und ihre Entwicklung* [Social inhibition and its development]. Berlin: Springer-Verlag.

Asendorpf, J. B. (1990a). The expression of shyness and embarrassment. In R. Crozier (Ed.), *Shyness and embarrassment: Perspectives from social psychology* (pp. 87–118). Cambridge: Cambridge University Press.

Asendorpf, J. B. (1990b). Development of inhibition during childhood: Evidence for situational specificity and a two-factor model. *Developmental Psychology*, **26**, 721–730.

Asendorpf, J. B., & Wallbott, H. G. (1982). Contributions of the German "Expression Psychology" to nonverbal communication research: Part I. Theories and concepts. *Journal of Nonverbal Behavior*, **6**, 135–147.

Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development from infancy. *Child Development*, **57**, 251–274.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, **62**, 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality, and self-opinion. *Personality and Individual Differences*, **5**, 11–25.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, **52**, 281–302.

Daly, S. (1978). Behavioural correlates of social anxiety. *British Journal of Social and Clinical Psychology*, **17**, 117–120.

Duncan, S. D., Jr., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. Hillsdale, NJ: Lawrence Erlbaum.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of Communication*, **22**, 353–374.

Emmerich, W. (1964). Continuity and stability in early social development. *Child Development*, **35**, 311–332.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgments. *Journal of Personality and Social Psychology*, **55**, 149–158.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging*, **1**, 159–171.

Huesmann, L. R., Eron, L. D., Lefkowitz, M. M., & Walder, L. O. (1984). Stability of aggression over time and generations. *Developmental Psychology*, **20**, 1120–1134.

Jones, W. H., Briggs, S. R., & Smith, T. G. (1986). Shyness: Conceptualization and measurement. *Journal of Personality and Social Psychology*, **51**, 629–639.

Jones, W. H., Cheek, J. M., & Briggs, S. R. (Eds.). (1986). *Shyness: Perspectives on research and treatment*. New York: Plenum Press.

Kagan, J. (1980). Perspectives on continuity. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 26–74). Cambridge, MA: Harvard University Press.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development*, **58**, 1459–1473.

Kopp, C. B., & McCall, R. B. (1980). Stability and instability in mental performance among normal, at-risk and handicapped infants and children. In P. B. Baltes & O. G. Brim (Eds.), *Life-span development and behavior* (Vol. 4, pp. 33–61). New York: Academic Press.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, **55**, 95–107.

Moss, H. A., & Susman, E. J. (1980). Longitudinal study of personality development. In O. G. Brim, Jr., & J. Kagan (Eds.), *Constancy and change in human development* (pp. 530–595). Cambridge, MA: Harvard University Press.

Petrinovich, L. (1979). Probabilistic functionalism: A conception of research method. *American Psychologist*, **34**, 373–390.

Rose, S. A., Feldman, J. F., & Wallace, I. F. (1988). Individual differences in infants' information processing: Reliability, stability, and prediction. *Child Development*, **59**, 1177–1197.

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, **8**, 467–487.

Weinert, F. E., & Schneider, W. (Eds.). (1989). *The Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC): No. 6. Longitudinal findings for wave 1–3* (res. rep.). Munich, Germany: Max-Planck-Institut.

Wilson, R. S. (1983). The Louisville twin study: Developmental synchronies in behavior. *Child Development*, **54**, 298–316.

Zivin, G. (1985). *The development of expressive behavior*. Orlando: Academic Press.

Manuscript received January 1, 1990; revised July 5, 1991.

Copyright © 2002 EBSCO Publishing