# Moderators of Self–Other Agreement: Reconsidering Temporal Stability in Personality

Jeremy C. Biesanz and Stephen G. West
Arizona State University

William G. Graziano
Texas A&M University

Accurate prediction requires information not only about central tendencies but also about variability. In personality prediction, however, most research has focused on trait-level central tendencies. Previously proposed moderators of personality prediction are all conceptually similar in comparing an individual's central tendency in response patterns with that of the normative person. This article proposes an alternative: Trait-level prediction is enhanced by measuring the temporal stability of response patterns within persons. Across 2 studies, individuals with temporally stable response patterns had higher self–other agreement on conscientiousness and extraversion than did individuals with less temporally stable patterns. By comparison, normatively based variables (interitem variability, scalability, or construct similarity) did not moderate self–other agreement. The implications for personality structure, assessment, and prediction are discussed.

A fundamental task of personality research is to determine how to conceptualize and measure personality so as to better understand and predict behavior. During the 1970s, personality researchers undertook a major reassessment of the relationship between traits and behavior. Dissatisfied with what Mischel (1968) had identified as an apparent correlational ceiling of .30, researchers took several different approaches to understanding and explaining these empirical relationships (West, 1983). Bem and Allen (1974) hypothesized that low correlations between individuals and behavior could occur if personality influences behavior only for some individuals. Bem and Allen argued that combining individuals who do with those who do not have a particular trait in analyses would attenuate the resulting trait–behavior correlations.

This moderator variable approach stands in distinct contrast to nomothetic trait-based approaches to personality, which, in their simplest form, assume that all individuals can be characterized as possessing the particular trait under consideration. In this view, the central question becomes the assessment of the individual's position on the underlying trait dimension. By reintroducing Allport's (1937) idea that perhaps not all traits are equally relevant to all individuals, researchers then face a second assessment task. After assessing each individual's position on the underlying trait dimension, researchers must assess traitedness: how strongly, if at all, that trait influences the individual's behavior.

The moderator variable strategy has sparked considerable re-

search. To date, there is substantial evidence that self-reports of *consistency* in behavior across situations, *relevance* of personality dimensions to behavior, and the *uniqueness* of how a personality dimension is behaviorally expressed are able to moderate self–other agreement (Bem and Allen, 1974; Cheek, 1982; Kenrick & Stringfield, 1980; Mischel & Peake, 1982; Zuckerman, Bernieri, Koestner, & Rosenthal, 1989; Zuckerman et al., 1988; Zuckerman, Miyake, Koestner, Baldwin, & Osborne, 1991; but see Chaplin & Goldberg, 1984, for a failure to replicate). Research has focused on self–other agreement as ratings by knowledgeable informant reflect an implicit pooling of observations of an individual's behavior across different situations and time (Epstein, 1983; Graziano, Jensen-Campbell, & Finch, 1997; Kenrick & Braver, 1982).

These moderators, which are highly dependent on individuals' inferences about their own past behavior, have several deficiencies, including low test–retest reliability (see Amelang & Borkenau, 1986). Consequently, a separate body of literature developed moderators that were less dependent on the quality of the participants' own inferential processes. Bem and Allen (1974) examined participant item response patterns on a conscientiousness scale as a potential moderator of self–other agreement on conscientiousness. They created an ipsatized measure in which each participant's variability between items on conscientiousness, relative to that participant's interitem variability on other scales, was used as a moderator variable. Bem and Allen reasoned that participants with low relative interitem variance would be more conscientious cross-situationally than participants with high relative interitem variance. Consistent with their prediction, Bem and Allen found that individuals with low relative interitem variance had higher self–other agreement than individuals with high relative interitem variance.

The use of relative interitem variability has been criticized, however, by several authors. These authors argued that using relative variability introduces additional error into the moderator; the use of raw variability has also been criticized because it loses differences in situational evocativeness (Lanning, 1988;

Paunonen & Jackson, 1985; Tellegen, 1988). Consequently, more recent research has not ipsatized interitem variability (Baumeister, 1991; Baumeister & Tice, 1988) but has standardized responses before computing interitem variability to minimize these problems (Britt, 1993).

Besides interitem variability, other moderators have been proposed to extract useful information from scale-response patterns. Lanning's (1988) scalability and Chaplin's (1991) construct similarity are two more recently proposed moderators calculated on an individual's response pattern. Scalability is a function of the sum of differences between a person's response profile and his or her predicted profile given the normative response profile. Construct similarity is the $Q$ correlation (see Stephenson, 1952) between a person's response profile and the normative response profile. The same theoretical prediction is made for each of these three moderators: Higher trait–behavior correlations are predicted for individuals whose response pattern mirrors, compared with those individuals whose response pattern does not mirror, the normative response pattern. The central difference among these three normative moderators is in the method of calculation of the measure of fit between an individual's response pattern and the normative response pattern.

Given the strong epistemological and measurement similarities among these three proposed moderators, one would expect high correlations among these moderators, as well as a similar ability to moderate relations with external criteria. These moderators have shown promise in predicting the magnitude of self–other agreement in some studies (Chaplin, 1991; Mischel & Peake, 1982), but other studies have failed to replicate the findings (Chaplin & Goldberg, 1984; Hofstee & Smid, 1986), and still other studies have yielded inconsistent results across moderators (Chaplin, 1991; Lanning, 1988).

## Temporally Stable Response Patterns

Inherent in the conceptualization of moderator variables in personality is a fundamental idea: If two individuals are at the same location on the latent trait (e.g., have equal levels of conscientiousness) but differ on traitedness, the more traited individual will demonstrate more consistent behavior with respect to that trait than the less traited individual. We treat traitedness as a continuous variable, varying in degree, in contrast to some earlier conceptions (e.g., Baumeister & Tice, 1988), which treated individuals as dichotomously either traited or untraited on a particular personality dimension.

If we can identify individuals who demonstrate consistent behaviors, then those individuals are traited and will show a stronger link between the personality trait and behavior. To date, consistency has been operationalized as nomothetic cross-situational consistency. We break with this tradition, using another form of consistency as an indicator of the degree of traitedness. As Ozer (1986) has noted, there are several different ways to conceive of consistency in addition to nomothetic cross-situational consistency. Allport (1937) suggested an important alternative form of consistency in his commentary on Hartshorne and May's (1928) study that found low cross-situational correlations for honesty in children. Allport (1937) wrote that these low correlations "prove only that children are not consistent *in the same way*, not that they are inconsistent with *themselves*"

(p. 250). If we examine individuals' scale response patterns assessed at a single measurement occasion and ask whether they are consistent with themselves, the answer is that we do not know—there are insufficient data to make this judgment. Researchers must know whether individuals always behave in the same manner within these situations over time. Sampling behavior across items does not provide an estimate of the consistency of individuals within themselves. Behavior must be sampled both across items and over time.

This reasoning suggests that instead of comparing individuals' scale response pattern with that of the average person, it may be more useful to compare each person's current response pattern with his or her pattern collected on other measurement occasions. Regardless of the shape of the response pattern, individuals who are consistent in their response patterns across time should be more predictable with respect to that trait, thus leading to higher self–other agreement.

In the present study we compared the normative moderators interitem variability, scalability, and construct similarity with a new moderator variable, temporal response pattern stability, using two broad traits, conscientiousness and extraversion. Conscientiousness has been the focus of much of the research on moderators, and initial results indicated that the normative moderators were successful in predicting the degree of self–other agreement (e.g., Bem & Allen, 1974; Mischel & Peake, 1982). Extraversion is a second broad trait from the five-factor structural approach to personality that has been the focus of considerable research attention (e.g., Watson & Clark, 1997). Appreciable levels of consensus and self–other agreement have been found for both traits, even when the observers have minimal information about the target person (Borkenau & Liebler, 1993; Kenny, Horner, Kashy, & Chu, 1992). The specific goals of the present study were (a) to test temporal response pattern stability as a moderator of self–other agreement, (b) to provide an opportunity to replicate previous findings for the normative moderators, and (c) to compare the relationships among these moderators.

## Study 1

### Method

#### Participants

Introductory psychology students ($N = 134$) were recruited to participate in return for partial fulfillment of their introductory psychology class requirements. A total of 111 participants completed the basic study requirements of attending three measurement sessions. Participants were also encouraged to bring two acquaintances into the laboratory in exchange for extra credit toward fulfillment of their course requirements. In addition, participants provided consent for us to obtain, via mail, a parental rating. Following Epstein (1983), we included only participants who had at least two complete informant ratings in the moderator analyses. This left a final study sample size of 102 (54 women and 48 men; mean age $= 19.68$ years, $SD = 2.29$) with ratings from two informants. Seventy-nine of these were rated by three informants.

#### Materials

Participants, peers, and parents rated the participant on 19 unipolar trait adjectives on conscientiousness and 20 unipolar trait adjectives on

extraversion (Goldberg, 1992), which were embedded within unrelated trait adjectives. The adjective *haphazard* was dropped from consideration for conscientiousness as several participants asked for a definition (see Graziano, Jensen-Campbell, Steele, & Hair, in press, regarding unknown words in standard personality assessment). All ratings were on a 9-point scale ranging from 0 (*extremely inaccurate*) to 8 (*extremely accurate*). Participant's self-rating instructions were modified from Goldberg (1992) to limit self-assessments of behavior to the previous week.[1] This change encouraged reporting of more variability in the trait adjectives over time as opposed to general or typical behavior. The specific rating instructions were as follows:

> *Participant Trait Rating Instructions*
> Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you were *during this past week*, as compared to other persons you know of the same sex and roughly your same age. Before each trait, please write a number indicating how accurately that trait describes you *for the past week*.

Peers and parents received Goldberg's (1992) standard rating instructions with the participant's name embedded within the instructions and used the same rating scale as participants. The specific rating instructions were as follows:

> *Peer and Parent Rating Instructions*
> Please use this list of common human traits to describe [participant's name] as accurately as possible. Describe [participant's name] as you see [him/her] at the present time, as compared to other persons you know of the same sex and roughly the same age. Before each trait, please write a number indicating how accurately that trait describes [participant's name].

## Design and Procedure

Participants completed the self-report inventory three times, at no less than 1-week intervals, in a lecture hall reserved for that purpose (mean interval between first and third assessment was 16.00 days, *SD* = 6.27). Peers were separated from the participant they rated, and when two peers rated the same participant simultaneously, they were separated from each other. Questionnaires were mailed to the parent designated by participants. Peer questionnaires were prefaced with explanatory instructions; parent questionnaires were prefaced with the same instructions in the form of an explanatory cover letter.

## Moderator Calculations

Values for each of the four moderator variables were calculated separately for extraversion and conscientiousness. Adjectives were reverse-coded when appropriate so that higher scores always indicated greater conscientiousness or extraversion. Reverse coding was performed before the values of the moderator variables were calculated.

*Temporal response pattern stability.* To quantify the temporal stability of response patterns, we computed the correlation between the pattern of responses for each pair of assessments within each participant within each trait after subtracting the normative response profile from each assessment. This correction reduced the impact of normative differences between adjectives and allows for a more straightforward interpretation of the resulting correlation.[2] The correlation between the two profiles provided an estimate of the similarity of the pattern of responses between these two assessments that was independent of mean response level (Cronbach & Gleser, 1953; see also Colvin, 1993, for a similar example). The three pairwise profile correlations (i.e., Time 1 and Time 2,

Time 1 and Time 3, and Time 2 and Time 3) were averaged into a single composite measure of temporal response pattern stability. Generalizability across the three measures of temporal response pattern stability was adequate ($\rho^2$ = .60 for conscientiousness and $\rho^2$ = .71 for extraversion; see Shavelson & Webb, 1991).

*Scalability, interitem variability, and construct similarity.* A mean response profile was computed for each individual by averaging their responses within adjectives across the three administrations. The normative moderators were computed on this mean profile for each individual.

Lanning's (1988) scalability index was computed according to the following formula: Scalability$i$ = $-\Sigma |X_{ij} - (X_{i\bullet} + X_{\bullet j} - X_{\bullet\bullet})|$.

Person *i*'s scalability was a function of his or her response to a single adjective ($X_{ij}$), person *i*'s mean level across adjectives ($X_{i\bullet}$), the average response in the full sample to that adjective ($X_{\bullet j}$), and the grand mean across participants and adjectives ($X_{\bullet\bullet}$). By multiplying the sum by negative one, we coded scalability such that individuals with higher values were closer to the mean response profile.

We computed interitem variability by taking the standard deviation across a participant's responses after standardization. Participant responses on item *j* were converted to *z* scores using the mean and standard deviation from the full sample. Following the procedure used in recent studies (Baumeister, 1991; Baumeister & Tice, 1988; Britt, 1993; Chaplin, 1991), we used the standard deviation as opposed to the variance because it results in a less skewed distribution.

We computed construct similarity (Chaplin, 1991) for each participant by calculating the *Q* correlation between the participant's response profile and the normative response profile.

## Results

### Attrition

Twenty-three of the 134 original participants completed only one or two assessments and thus did not fulfill the study requirements. Participants who failed to complete the study requirements did not differ from participants who completed the study requirements in gender, $\chi^2(1, N = 134) = .31, n s$; age, $t(124) = -.52, n s$; or extraversion, $t(131) = .72, n s$. However, participants who did not fulfill the study requirements had a significantly lower mean conscientiousness level than those participants who completed the study ($M$ = 4.56 vs. $M$ = 5.13), $d = .43, t(131) = 2.47, p < .02$.

### Moderator Intercorrelations

Table 1 presents the means, standard deviations, and correlations among the four moderator variables for conscientiousness and extraversion. The measures of scalability and interitem variability were significantly correlated within each trait. However, construct similarity was unassociated with all other moderators. Temporal response pattern stability was associated with lower

---

[1] The ratings obtained using the modified instructions were compared with those obtained using the original Goldberg (1992) instructions within a separate pilot study (*n* = 34). The one-week test–retest correlation comparing the two different instructions (order of instructions counterbalanced across participants) was substantial for both conscientiousness, $r(32) = .73, p < .01$, and extraversion, $r(32) = .83, p < .01$.

[2] Reported results do not substantially change if temporal response pattern stability is computed on either raw, uncorrected profiles or on fully standardized profiles (see Shoda, Mischel, & Wright, 1994, for an example of temporal stability on standardized profiles).

Table 1

*Moderator Variable Intercorrelations and Means Within Conscientiousness and Extraversion for Study 1*

| Moderator variable | Moderator | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. Temporal response pattern stability | .28** | .56** | −.62** | −.09 |
| 2. Interitem variability | .47** | .24* | −.95** | −.14 |
| 3. Scalability | −.43** | −.93** | .13 | .11 |
| 4. Construct similarity | −.18 | −.13 | .04 | .08 |
| Conscientiousness *M* (*SD*) | .28 (.20) | .70 (.21) | −13.78 (3.99) | .35 (.20) |
| Extraversion *M* (*SD*) | .37 (.23) | .77 (.22) | −16.27 (4.88) | .40 (.24) |

*Note.* Correlations above the diagonal are the moderator intercorrelations for extraversion. Correlations below the diagonal are moderator intercorrelations for conscientiousness. Correlations on the diagonal are the correlations between traits for the moderator variable.
$*p < .05.$  $**p < .01.$

scalability and higher interitem variability but was not significantly associated with construct similarity. Both temporal response pattern stability and interitem variability were associated across the traits extraversion and conscientiousness.

## Self–Other Agreement Moderation

The correlation between the mean of the self-reports across the three measurement occasions and the mean of the informant reports was significant for both conscientiousness, $r(100) = .34, p < .01$, and extraversion, $r(98) = .55, p < .01$. To examine possible moderation of self–other agreement, we performed multiple regression analyses separately for each moderator variable following procedures described by Aiken and West (1991). The mean of peer and parent ratings of participants served as the criterion. Generalizability among these three informant ratings was adequate ($\rho^2 = .64$ for conscientiousness and $\rho^2 = .76$ for extraversion). Predictor variables were centered (put in deviation score form) prior to analysis to facilitate interpretation of lower order terms in the equation (see Aiken & West, 1991). The predictors in the regression equation consisted of participant self-reported conscientiousness ($\rho^2 = .85$) or extraversion ($\rho^2 = .81$) averaged across the three administrations, the moderator variable (interitem variability, scalability, construct similarity, or temporal pattern stability), and the Self-Reported Trait Level × Moderator Variable product term.

As hypothesized, the interaction between temporally stable response patterns and self-reported trait level was a significant predictor of others' ratings for both conscientiousness, $F(1, 98) = 4.17, p < .05$, and extraversion, $F(1, 96) = 4.90, p < .03$. Participants with more temporally stable response patterns showed higher self–other agreement than participants with less temporally stable response patterns for both conscientiousness (see Figure 1) and extraversion (see Figure 2).

To probe the nature of the interactions, we conducted simple slope tests following the procedures outlined by Aiken and West (1991). We examined the simple slopes at low levels of temporal response pattern stability (one standard deviation below the mean) and high levels of temporal response pattern stability (one standard deviation above the mean) within each trait. For

conscientiousness, self–other agreement was not significantly different from zero at low levels of temporal response pattern stability, $t(98) = .61, ns$, but was significant at high levels of temporal response pattern stability, $t(98) = 3.82, p < .001$. For extraversion, self–other agreement was significantly different from zero at low levels, $t(96) = 2.84, p < .01$, as well as high levels of temporal response pattern stability, $t(96) = 6.00, p < .001$.

There was no evidence that the previously proposed normative moderators (interitem variability, construct similarity, and scalability) significantly moderated self–other agreement for either conscientiousness or extraversion (all $Fs < 2.80, ns$). Following Cohen (1988) and Aiken and West (1991), Table 2 reports a measure of effect size (partial correlation of the product term) for all interaction terms.

## Discussion

As predicted, individuals who displayed more stable response patterns across the three assessments had higher self–other
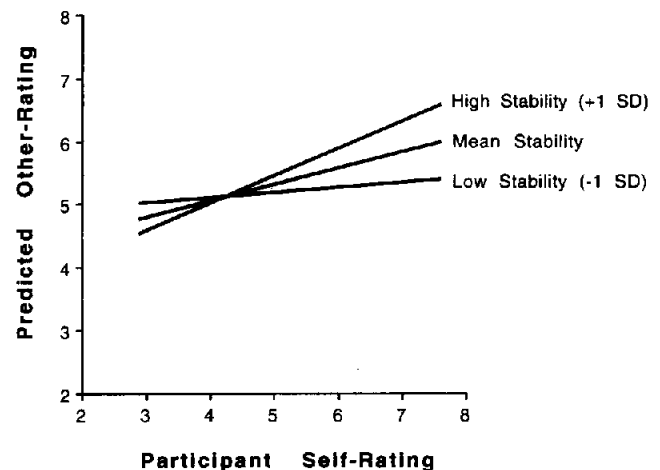


*Figure 1.* Predicted mean informant conscientiousness rating as a function of participant conscientiousness and temporal response pattern stability in Study 1.
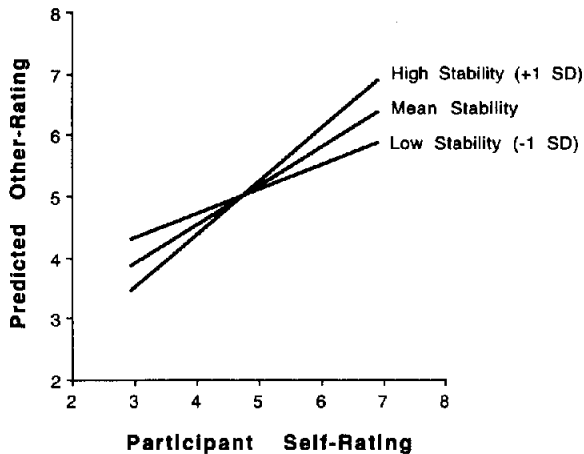
Figure 2. Predicted mean informant extraversion rating as a function of participant extraversion and temporal response pattern stability in Study 1.

agreement than individuals with less temporally stable response patterns. Temporally stable response patterns, moreover, were associated with higher variability across adjectives. In contrast to some previous studies, there was no evidence of the moderating ability of interitem variability or scalability. Indeed, the effect for these two moderators was in the direction *opposite* from that predicted. The effect of construct similarity for both extraversion and conscientiousness, though nonsignificant, was similar to the average effect size across traits in previous research (Chaplin, 1991). Given the history of failures to replicate moderating influences (Chaplin & Goldberg, 1984; Hofstee & Smid, 1986; see also Epstein, 1983), conclusions based on a single study must be made with caution. Consequently, we decided to replicate the study to determine the reliability of outcomes across heterogeneities of samples and time.

## Study 2

### Method

#### Participants

Introductory psychology students ($N = 178$) were recruited to participate in return for partial fulfillment of their course requirements. A total of 164 participants completed the study requirement of attending three measurement sessions.

As in Study 1, participants were encouraged to bring two acquaintances into the laboratory in exchange for additional credit toward fulfillment of their course requirements. Participants provided consent for obtaining a parental rating via mail. Only participants who had at least two complete informant ratings were retained for the moderator analyses. This left a final study sample size of 148 (107 women and 41 men; mean age = 19.75 years, $SD = 2.99$). Of these, 113 participants were rated by three informants.

### Materials

Participants, peers, and parents again rated the participant on 19 unipolar trait adjectives on conscientiousness and 20 unipolar trait adjectives on extraversion (Goldberg, 1992), which were embedded within a list of trait adjectives unrelated to conscientiousness and extraversion. The scales, as well as instructions for participants and for peers and parents, were identical to those in the previous study.

### Design and Procedure

Participants completed the self-report inventory three times, at no less than 1-week intervals, in several small classrooms reserved for that purpose (mean interval between first and third assessment was 15.06 days, $SD = 3.98$). Peers completed their ratings during one of the participant's assessments. When two peers rated the same participant simultaneously, they were separated from the participant they were rating and from each other. Questionnaires were mailed to the parent designated by participants. Moderator variables were calculated using the procedures outlined in Study 1. Generalizability across the three measures of temporal response pattern stability again was adequate ($\rho^2 = .72$ for conscientiousness and $\rho^2 = .77$ for extraversion).

## Results

### Attrition

Of the 178 original participants, 14 completed only one or two assessments and thus did not fulfill the study requirements. Participants who failed to complete the study requirements did not differ from participants who completed the study requirements in gender, $\chi^2(1, N = 178) = .52$, $ns$; age, $t(146) = .18$, $ns$; or extraversion, $t(173) = 1.42$, $ns$. However, as in Study 1, participants who did not fulfill the study requirements had a significantly lower mean conscientiousness level than those participants who did complete the study ($M = 4.56$ vs. $M =$

Table 2
*Moderator Effect Sizes for Participant–Other Agreement for Study 1*

| | Trait | |
|---|---|---|
| Moderator variable | Conscientiousness | Extraversion |
| 1. Temporal response pattern stability | .20* | .22* |
| 2. Interitem variability | .09 | .14 |
| 3. Scalability | −.08 | −.17† |
| 4. Construct similarity | .08 | .10 |

*Note.* Effect size is the partial correlation of the interaction of the moderator variable and the broad trait with informant ratings on that trait (Cohen, 1988).
† $p < .10$ (marginally significant). * $p < .05$.

5.42), $d = .45$, $t(173) = 2.93$, $p < .01$. Participants in Study 2, which was conducted at a much earlier point in a following semester, were, on average, more conscientious than participants in Study 1 ($M = 5.39$ vs. $M = 5.07$), $d = .34$, $t(257) = 2.70$, $p < .01$.

## Moderator Intercorrelations

The means, standard deviations, and correlations among the four moderators are presented in Table 3, and the results were consistent with those of Study 1. Measures of scalability and interitem variability were again significantly correlated. Temporal response pattern stability was again significantly associated with lower scalability and higher interitem variability. In contrast to the previous study, temporal response pattern stability was significantly correlated with lower construct similarity. As originally predicted, construct similarity was negatively correlated with interitem variability and positively correlated with scalability, resulting in significant intercorrelations among the normative moderators.

## Self–Other Agreement Moderation

The correlation between the mean of the self-reports across the three measurement occasions and the mean of the informant reports was significant for conscientiousness, $r(146) = .52$, $p < .01$, and extraversion, $r(145) = .52$, $p < .01$. To examine moderation of self–other agreement, we performed multiple regression analyses for each moderator following the procedures described by Aiken and West (1991). Once again, all predictor variables were centered prior to analysis. Peer and parent ratings of participants were averaged and served as the criterion. Generalizability across these three informant ratings was adequate ($\rho^2 = .64$ for conscientiousness and $\rho^2 = .71$ for extraversion). Participant self-reported conscientiousness ($\rho^2 = .89$) or extraversion ($\rho^2 = .91$) averaged across the three administrations, the moderator variable (interitem variability, scalability, construct similarity, or temporal pattern stability), and the Participant Rating × Moderator product term were entered into the regression equation.

The interaction between temporal response pattern stability and self-reported trait level was significant in predicting other ratings for conscientiousness, replicating the previous study, $F(1, 144) = 4.90$, $p < .03$, and was marginally significant for extraversion, $F(1, 143) = 3.33$, $p < .07$. Participants with more stable response patterns again showed higher self–other agreement than participants with less temporally stable response patterns for both conscientiousness and extraversion. The form of these interactions paralleled those of Study 1.

To explore further the nature of the interactions, we again conducted simple slope tests following the procedures suggested by Aiken and West (1991) at levels of low temporal response pattern stability (one standard deviation below the mean) and high levels of temporal response pattern stability (one standard deviation above the mean) within each trait. For conscientiousness, self–other agreement was significant at low levels of temporal response pattern stability, $t(144) = 2.87$, $p < .01$, as well as high levels of temporal response pattern stability, $t(144) = 6.68$, $p < .001$. For extraversion, self–other agreement was again significantly different from zero for low levels of temporal response pattern stability, $t(143) = 2.63$, $p < .01$, as well as high, $t(143) = 6.67$, $p < .001$.

The normative moderators (interitem variability, construct similarity, and scalability) did not significantly moderate self–other agreement for conscientiousness or extraversion (all $Fs < 1.10$, $ns$). Effect sizes (partial correlations of the product term) for all interaction terms are presented in Table 4.

## Effect Sizes Across the Two Studies

To obtain a better estimate of the effect of each moderator, we combined moderator effect sizes meta-analytically across the two studies within each trait following procedures suggested by Shadish and Haddock (1994). Combining across both studies, temporal response pattern stability significantly moderated the level of self–other agreement for both conscientiousness ($z = 2.96$, $p < .01$) and extraversion ($z = 2.77$, $p < .01$; see Table

Table 3

*Moderator Variable Intercorrelations and Means Within Conscientiousness and Extraversion for Study 2*

|  | Moderator | | | |
|---|---|---|---|---|
| Moderator variable | 1 | 2 | 3 | 4 |
| 1. Temporal response pattern stability | .27** | .57** | −.58** | −.32** |
| 2. Interitem variability | .59** | .23* | −.96** | −.16 |
| 3. Scalability | −.59** | −.96** | .17* | .18* |
| 4. Construct similarity | −.18* | −.19* | .14 | −.08 |
| Conscientiousness *M (SD)* | .35 (.24) | .70 (.23) | −14.96 (5.36) | .29 (.26) |
| Extraversion *M (SD)* | .41 (.24) | .74 (.22) | −17.61 (5.75) | .41 (.23) |

*Note.* Correlations above the diagonal are the moderator intercorrelations for extraversion. Correlations below the diagonal are moderator intercorrelations for conscientiousness. Correlations on the diagonal are the correlations between traits for the moderator variable.
* $p < .05$.  ** $p < .01$.

Table 4
*Moderator Effect Sizes for Participant–Other Agreement for Study 2*

| Moderator variable | Trait | |
|---|---|---|
| | Conscientiousness | Extraversion |
| 1. Temporal response pattern stability | .18* | .15† |
| 2. Interitem variability | .05 | .00 |
| 3. Scalability | −.09 | −.03 |
| 4. Construct similarity | .01 | .02 |

*Note.* Effect size is the partial correlation of the interaction of the moderator variable and the broad trait with informant ratings on that trait (Cohen, 1988).
† $p < .10$ (marginally significant).  * $p < .05$.

5).[3] The normative moderators did not significantly moderate self–other agreement for either conscientiousness or extraversion (all zs < 1.32, n s).

### Temporal Response Pattern Stability and Interitem Variability

For ease of presentation to this point, we have defined temporal response pattern stability as the mean of the Week 1 to Week 2, Week 1 to Week 3, and Week 2 to Week 3 Q correlations for each broad trait. However, this index suffers from a potential problem of interpretation, identified by Shoda, Mischel, and Wright (1994) in another context. Statistically, it is theoretically expected that temporal response pattern stability would be positively associated with interitem variability. Empirically, this was confirmed in both Study 1 and Study 2: Individuals who had more variability in their responses tended to have more temporally stable patterns. Interitem variability by itself did not moderate significantly self–other agreement, but it is possible that temporal response pattern stability is psychologically meaningful only in the context of substantial variability between response classes. Alternatively, the measurement of temporal response pattern stability may simply be attenuated for individuals with little variability between response classes. Therefore, a final set of analyses was performed to isolate further the effect of temporal pattern response stability.

In the first analysis, we removed statistically the influence of interitem variability from our measure of temporal response pattern stability within each trait. To do this, we entered the predictors temporal response pattern stability, interitem variability, self-report (conscientiousness or extraversion), and the Temporal Response Pattern Stability × Self-Report interaction (the moderator effect of interest) into a regression equation predicting other ratings. Across both studies for both conscientiousness and extraversion, the estimates of the unique effect of temporal response pattern stability as a moderator of self–other agreement were similar to those found in our original analyses reported above. For conscientiousness, temporal response pattern stability significantly moderated self–other agreement after meta-analytically combining across both studies ($r = .19$, $z = 2.84$, $p < .01$). For extraversion, temporal response pattern stability again significantly moderated self–other agreement after meta-analytically combining across both studies ($r = .18$, $z = 2.71$, $p < .01$).

In the second analysis, we entered the predictors temporal response pattern stability, interitem variability, self-report (conscientious or extraversion), the Interitem Variability × Self-Report interaction, and the Temporal Response Pattern Stability × Self-Report interaction (the moderator effect of interest) into a regression equation predicting other ratings. Once again, temporal response pattern stability significantly moderated self–other agreement above and beyond interitem variability and its interaction with trait level, after meta-analytically combining across both studies, for both conscientiousness ($r = .17$, $z = 2.76$, $p < .01$) and extraversion ($r = .18$, $z = 2.84$, $p < .01$).

Given the lack of any appreciable difference in effect sizes between the original uncorrected and the subsequent analyses, which used two different corrections for the effects of interitem variability, it appears that interitem variability cannot account for the present findings. Nonetheless, individuals with substantial interitem variability can more easily obtain higher temporal response pattern stability coefficients than individuals with lower interitem variability given similar levels of within-adjective temporal instability.[4]

### General Discussion

The results of Study 2 are consistent with those of Study 1. Across both studies, participants with more temporally stable response patterns had higher self–other agreement for extraversion and conscientiousness than participants with less temporally stable response patterns. The replicated findings for conscientiousness are especially encouraging given the differences in mean levels of conscientiousness between Study 1, which was

[3] Whether temporal response pattern stability is calculated on the raw scores (uncorrected for normative differences between items) or on fully standardized profiles (corrected for both normative mean and variability differences between items), the results are parallel to those presented here. Across both studies, using raw profiles to compute temporal response pattern stability yields a moderator effect size for conscientiousness of $r = .20$, $z = 3.32$, $p < .001$, and for extraversion of $r = .18$, $z = 2.96$, $p < .01$. If profiles are standardized before computing temporal response pattern stability, the moderator effect size estimate, across both studies, for conscientiousness is $r = .17$, $z = 2.72$, $p < .01$, and for extraversion is $r = .18$, $z = 2.87$, $p < .01$.

[4] We are grateful to an anonymous reviewer for suggesting these analyses.

Table 5
*Moderator Effect Sizes for Participant–Other Agreement Combined Across Studies*

| | Trait | |
|---|---|---|
| Moderator variable | Conscientiousness | Extraversion |
| 1. Temporal response pattern stability | .19** | .18** |
| 2. Interitem variability | .07 | .06 |
| 3. Scalability | −.08 | −.09 |
| 4. Construct similarity | .03 | .05 |

*Note.* Effect size is the partial correlation of the interaction of the moderator variable and the broad trait with informant ratings on that trait.
** $p < .01$.

conducted late in the semester, and Study 2, which was conducted earlier in a subsequent semester. The suggestive evidence found in Study 1 for construct similarity was not replicated, however.

Taken together, these two studies illustrate the limitations of the information available on conscientiousness and extraversion with only one scale administration. Each individual's responses potentially contain valuable information in addition to his or her average response. Much of this information, however, is not captured solely by the match between the normative response pattern and the participant's response pattern at any given assessment. Instead, it is the stability of that pattern of responses across measurement occasions that yields a better understanding of how others view each participant.

We did not find support across our two studies for the normative moderators interitem variability, scalability, and construct similarity.[5] Bem and Allen (1974), Mischel and Peake (1982), and Chaplin and Goldberg (1984) found support for ipsatized interitem variance as a moderator of self–other agreement for conscientiousness. Ipsatizing introduces information about the differences among the mean trait levels and the other measured traits into this measure (see Tellegen, 1988). The choice of traits to measure other than conscientiousness may be crucial to the success of ipsatized interitem variance in moderating self–other agreement for conscientiousness. Interpreting the effect is even more difficult: Is the effect due to interitem variability within conscientiousness, interitem variability within other traits, or mean trait-level differences?

Across studies that examined scale-response pattern moderators other than ipsatized interitem variability, the results for conscientiousness are decidedly mixed. Chaplin (1991) found support for scalability and interitem variability, but not construct similarity. Lanning (1988) found significant support for scalability as a moderator in only one in six regression analyses. In the only study of which we are aware to examine extraversion, Hofstee and Smid (1986) did not find support for interitem variability as a moderator of self–other agreement across any of the 14 traits that they assessed. In light of the present and past research, the status of these normative moderators of self–other agreement appears tenuous.

*Toward Understanding the Processes Underlying Moderation*

On the basis of previous theorizing and research, we identified two explanations that could contribute to an understanding of

our finding that temporal pattern response stability moderates self–other agreement. These explanations appear interrelated but reflect different levels and modes of theorizing.

The first explanation involves the concept of metatraits (Baumeister, 1991; Baumeister & Tice, 1988). In this perspective, traited individuals possess the trait in question and can be expected to show substantial levels of self–other agreement, whereas untraited individuals do not possess the trait and would be expected to show no agreement with another on their level of that specific trait. If we assume that individuals who are one standard deviation below the mean in temporal response pattern stability are untraited, then we would expect that the slope of the regression line would be zero at this point. Our data, however, were inconsistent with the metatrait hypothesis. In three of the four tests of simple slopes at this low value of temporal response pattern stability within our regression analyses, as well as in exploratory statistical analyses using slicing techniques (see Cook & Weisberg, 1994), outcomes did not support the metatrait prediction. Taken together, the outcomes do not corroborate the strong form of the metatrait explanation, which implies a qualitative difference between traited and untraited individuals.

There is a second explanation, derived in part from research and theorizing by Epstein (1979, 1980, 1983) and Mischel and Shoda (1995). There may be individual differences in the reliability of trait–behavior relationships without the strong presumption that there are some untraited individuals who have no trait-level reliability. Individuals with low temporal pattern stability may show greater variability in the situations in which they participate on a regular (e.g., weekly) basis or may be more responsive to the influences of the situations on their behavior. Consequently, when asked to report on their behavior for a given week, they will, on average, be more discrepant from their "true" trait level than individuals with more temporally stable patterns. Reflecting this increased variability, reports by knowledgeable informants may show less agreement both across informants and over time. From the standpoint of prediction, if we aggregate reports over time and aggregate reports from multiple informants who observe the participants in different situations, then we may obtain self-reports and informant reports that are comparable in reliability to those from more temporally stable

---

[5] Construct similarity has substantial, but not complete overlap with the concept of shared meaning (see Chaplin & Panter, 1993; Kenny, 1991).

participants—under the present measurement conditions. Within conditions in which comparable levels of reliability are achieved in the high and low temporally stable individuals, comparable levels of self–other agreement would be expected.

## Implications for Personality Coherence

Shoda et al. (1994) demonstrated similar evidence of stable patterns of responses by children over time. They also indicated that there are substantial individual differences in the stability of these response patterns. Our study replicated these general findings. The mean temporal response pattern stability coefficients for both extraversion and conscientiousness in our study were substantial. At least some individuals showed unique and stable patterns of response across the different subdimensions or facets that constitute the broad traits of conscientiousness and extraversion. Most important, we found that knowledgeable informants more accurately predicted the mean trait level of students who have what Shoda et al. termed a stable "behavioral signature."

Our data differed from those of Shoda et al. (1994), however, in several important ways as a function of the measures, participants, and settings involved in each study. First, Shoda et al. studied the temporal stability of the profile of a single broad response class (e.g., verbal aggression) to a fixed set of psychological situations, whereas we studied the temporal stability of a set of response classes (responses to Goldberg's, 1992, trait adjectives) without specifying situational referents. These are two different perspectives and measures of personality coherence that consequently can offer different, and we hope complementary, insights into personality (see Ozer, 1986, for a more detailed theoretical discussion). Second, the college students in our study had surely progressed further in the development of their adult personality structures than the 10-year-old children in the Shoda et al. study. College participants may display patterns of responses that are more likely to be perceived as generally coherent even in the absence of specifying particular psychological situations (e.g., Graziano et al., in press). Third, the reports of our college participants, their peers, and parents were presumably based on behaviors of the participants in a wider variety of more freely chosen settings than the limited number of structured summer camp settings studied by Shoda et al. Yet we were able to find considerable evidence of the stability of the response profiles and substantial self–other agreement in judgments of the broad personality traits of conscientiousness and extraversion for participants with temporally stable response patterns. One possibility is that our informants may have inferred the participants' level on each of the facets of conscientiousness and extraversion from a fairly standard set of common psychological situations, reducing the importance of enumerating specific situations. However, that agreement was based on the ratings of peers and parents, who may not see the participants in similar psychological situations, weakens such an argument (see Kenny, 1991). Our findings may underscore the importance of the participants' ability to choose situations freely. At least in naive phenomenology, constrained choices provide less information regarding the individual as the preponderant cause of action than choices made freely (Ickes, Snyder, & Garcia, 1997; Jones & Davis, 1965).

In their recent theorizing, Mischel and Shoda (1995) focused on the psychological processes that may underlie stable and distinctive patterns of behavior across situations. In their formulation, the structure and invariance of personality is captured by individual differences in cognitive–affective units (i.e., expectations, affects, goals, and self-regulatory plans) that are activated when encoding situations. The complex and dynamic interaction among these cognitive–affective units results in distinct and stable behavioral patterns that can be characterized as if . . . then . . . situation–behavior profiles that "constitute the basic phenomena of personality" (Shoda et al., 1994, p. 684).

Our finding about individual differences in patterns of response classes may have implications for the Mischel and Shoda (1995) theoretical framework. This is best illustrated with an example. Consider two individuals, Jane and Jessica, who have equivalent levels of extraversion after aggregating across response classes. Jane may be generally more assertive and less active than most people, whereas Jessica may be precisely the opposite—more active but less assertive. Further, Jane may be very active and animated while lecturing and unassertive during committee meetings, whereas Jessica is fairly passive while lecturing but assertive during meetings. To more fully understand Jane and Jessica's personologically meaningful variability, one must account for not only how they interpret each situation (if . . . ) but also their natural proclivity for engaging in the suggested specific behavior (then . . .). An individual's distinctive pattern of behavior may therefore be a function of both which specific behavior is suggested in each situation and her or his general tendency or capability to engage in that specific behavior (Graziano et al., 1997, p. 393).

Mischel and Shoda (1995) noted that the challenge facing personality psychology is how to "conceptualize and demonstrate the type of behavioral coherence that is produced by the invariant qualities within the person" (p. 247). We concur and suggest that both the existence of meaningful and coherent patterns of behavior as well as the temporal stability of those behavioral patterns constitute important aspects of personality.

## Resolving the Bem and Allen (1974) Paradox

Bem and Allen (1974) posed the paradox that our intuitions tell us that there are pervasive cross-situational consistencies in behavior, yet research tells us otherwise. Psychology is replete with instances of mistaken intuitions; however, we cannot disregard our intuitions concerning consistency (Kenrick & Funder, 1988). Bem and Allen attempted to solve this paradox by identifying individuals who display cross-situational consistency. Yet, as we have noted previously, the data do not appear to be supportive of this proposed solution.

We propose another solution. Instead of asking whether people are consistent like other people, or even which people are consistent like other people, the more useful question may be which people are consistent with themselves over time. This solution is compatible with the Mischel and Peake's (1982) finding that the temporal stability of people's behavior is linked to their perceptions of their own consistency (for a related perspective on self-esteem, see Kernis & Waschull, 1995, pp. 97–107).

This within-person temporal-stability conception may reflect more closely how people form intuitions and impressions about the consistency of behavior in others. Imagine, for example, that a new male coworker just missed his flight to an important conference. If his female colleague is then asked to assess his conscientiousness and punctuality, clearly her impression and prediction would be enhanced if she knew he regularly missed his flight to important conferences. It is the temporal stability (or instability) of patterns of behavior that helps determine how people come to predict how others will behave. In all, the evidence presented here suggests that closer examination of and research into the organization and stability of variables within individuals may lead to a better understanding of how we perceive others and the relationship between personality and behavior.

## References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Amelang, M., & Borkenau, P. (1986). The trait concept: Current theoretical considerations, empirical facts, and implications for personality inventory construction. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 7–34). Berlin, Germany: Springer-Verlag.

Baumeister, R. F. (1991). On the stability of variability: Retest reliability of metatraits. *Personality and Social Psychology Bulletin, 17*, 633–639.

Baumeister, R. F., & Tice, D. M. (1988). Metatraits. *Journal of Personality, 56*, 571–598.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review, 81*, 506–520.

Borkenau, P., & Liebler, A. (1993). Consensus and self–other agreement for trait inferences from minimal information. *Journal of Personality, 61*, 477–496.

Britt, T. W. (1993). Metatraits: Evidence relevant to the validity of the construct and its implications. *Journal of Personality and Social Psychology, 65*, 554–562.

Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality, 59*, 143–178.

Chaplin, W. F., & Goldberg, L. R. (1984). A failure to replicate the Bem and Allen study of individual differences in cross-situational consistency. *Journal of Personality and Social Psychology, 47*, 1074–1090.

Chaplin, W. T., & Panter, A. T. (1993). Shared meaning and the convergence among observers' personality descriptions. *Journal of Personality, 61*, 553–585.

Cheek, J. M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer-rating study. *Journal of Personality and Social Psychology, 43*, 1254–1269.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colvin, C. R. (1993). Childhood antecedents of young-adult judgability. *Journal of Personality, 61*, 611–635.

Cook, R. D., & Weisberg, S. (1994). *An introduction to regression graphics*. New York: Wiley.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50*, 456–473.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*, 1097–1126.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35*, 790–806.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51*, 360–392.

Goldberg, L. R. (1992). The development of markers for the Big Five factor structure. *Psychological Assessment, 4*, 26–42.

Graziano, W. G., Jensen-Campbell, L. A., & Finch, J. F. (1997). The self as a mediator between personality and adjustment. *Journal of Personality and Social Psychology, 73*, 392–404.

Graziano, W. G., Jensen-Campbell, L. A., Steele, R. G., & Hair, E. C. (in press). Unknown words in self-reported personality: Lethargic and provincial in Texas. *Personality and Social Psychology Bulletin*.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan.

Hofstee, W. K. B., & Smid, N. G. (1986). Psychometric models for analysis of data from personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 166–177). Berlin, Germany: Springer-Verlag.

Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 165–195). San Diego, CA: Academic Press.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98*, 155–163.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88–97.

Kenrick, D. T., & Braver, S. L. (1982). Personality: Idiographic and nomothetic! A rejoinder. *Psychological Review, 89*, 182–186.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person–situation debate. *American Psychologist, 43*, 23–34.

Kenrick, D. T., & Stringfield, D. O. (1980). Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review, 87*, 88–104.

Kernis, M. H., & Waschull, S. B. (1995). The interactive roles of stability and level of self-esteem: Research and theory. In M. Zanna (Ed.), *Advances in experimental social psychology (Vol. 27*, pp. 93–141). San Diego, CA: Academic Press.

Lanning, K. (1988). Individual differences in scalability: An alternative conception of consistency for personality theory and measurement. *Journal of Personality and Social Psychology, 55*, 142–148.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Mischel, W., & Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review, 89*, 730–755.

Mischel, W., & Shoda, Y. (1995). A cognitive–affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*, 246–268.

Ozer, D. J. (1986). *Consistency in personality*. Berlin, Germany: Springer-Verlag.

Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review, 92*, 486–511.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–284). New York: Sage.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology, 67,* 674–687.

Stephenson, W. (1952). Some observations of the *Q* technique. *Psychological Bulletin, 49,* 483–498.

Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality, 56,* 621–663.

Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 767–793). San Diego, CA: Academic Press.

West, S. G. (Ed.). (1983). Personality and prediction: Nomothetic and idiographic approaches [Special issue]. *Journal of Personality, 51*(3).

Zuckerman, M., Bernieri, F., Koestner, R., & Rosenthal, R. (1989). To predict some of the people some of the time: In search of moderators. *Journal of Personality and Social Psychology, 57,* 279–293.

Zuckerman, M., Koestner, R., DeBoy, T., Garcia, T., Maresca, B. C., & Sartoris, J. M. (1988). To predict some of the people some of the time: A reexamination of the moderator variable approach in personality theory. *Journal of Personality and Social Psychology, 54,* 1006–1019.

Zuckerman, M., Miyake, K., Koestner, R., Baldwin, C. H., & Osborne, J. W. (1991). Uniqueness as a moderator of self–peer agreement. *Personality and Social Psychology Bulletin, 17,* 385–391.

---

## AMERICAN PSYCHOLOGICAL ASSOCIATION
## SUBSCRIPTION CLAIMS INFORMATION

Today's Date:_____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do NOT duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION

MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL)

ADDRESS

DATE YOUR ORDER WAS MAILED (OR PHONED)

____PREPAID ____CHECK ____CHARGE
CHECK/CARD CLEARED DATE:_____

CITY          STATE/COUNTRY          ZIP

(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.)

YOUR NAME AND PHONE NUMBER

ISSUES: ____ MISSING ____ DAMAGED

TITLE

VOLUME OR YEAR          NUMBER OR MONTH

*Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.*

━━━━━━━━ (TO BE FILLED OUT BY APA STAFF) ━━━━━━━━

DATE RECEIVED: _____          DATE OF ACTION: _____
ACTION TAKEN: _____          INV. NO. & DATE: _____
STAFF NAME: _____          LABEL NO. & DATE:_____

**Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242**

**PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.**