We are pleased to note that Bem (1977) does not question our basic results (Allen & Potkay, 1977). Once again we have demonstrated that individuals' self-descriptions vary markedly on a day-to-day basis (Allen & Potkay, 1973). In addition, this time we have shown that the intraindividual variance is controlled in large part by "significant events."

Accepting our results, Bem's primary objection to the research is that the Adjective Generation Technique (AGT) is a nonidiographic measure. Contrary to this opinion we believe the AGT to be a uniquely idiographic person-centered instrument. It relies entirely upon the individual's own phenomenology and construct system for adjective descriptions. It allows each person full opportunity to convey his or her own central set of predispositions, at the discretion of each person alone. The AGT is not "quasi-projective," it is fully projective. Subjects simply are asked to think of themselves at the end of a given day and to write down adjectives which best describe themselves. The openendedness of this instruction is clear and obvious. No cues are provided to subjects whatsoever.

Bem has argued that the basic "deficiency" of the AGT as an idiographic method is that we have had people other than our subjects rate the favorability of the words that our subjects generated. But does having someone other than our subjects rate adjectives render our technique only quasi-idiographic? Not necessarily. It may be that others' favorability ratings can be reasonably substituted for the subjects' own ratings just as persons' inferences

1. Thanks are due to Drs. Gene Smith and Eric Ward for their helpful comments on earlier drafts of this paper.
2. Requests for reprints should be sent to Bem P. Allen, Department of Psychology, Western Illinois University, Macomb, Illinois 61455.
Intraindividual variability: Reply to Bem

about the attitudes of a target can be substituted for the target’s own inferences (Bem, 1967). In beginning to investigate this possibility we had subjects generate adjectives to describe themselves and then rate the adjectives on our 7-point (0-6, where 6 is the most favorable possible) favorability scale (the list values are in the hundreds because we use 100 raters and then take total ratings). As always, subjects (or raters) were provided with no a priori definition of favorability. Results showed that 56 of 64 subjects made mean ratings whose values were within 1 scale point of the mean list values \( p < .0000001 \). The eight subjects for whom correspondence was not within 1 scale point assigned each of their five adjectives a 6 resulting in a mean of 6 for self-rated values. For these eight subjects correspondence could not have been within 1 scale point because only a few list values are as high as 5 and none are as high as 6. Thus our preliminary investigation indicates reasonable correspondence between favorability values assigned by subjects to their adjectives and values from our favorability list.

The above logic with regard to the reasonable correspondence of AGT favorability ratings also could be applied to the ratings of the favorability of events. Concerning the rating of favorability of events, Bem (1977) indicates that “blind judges” could not rate the event “started my period” accurately. We disagree, partly because the judges have more information than simply “started my period.” Persons reporting events typically make comments about the event, thus providing raters with information that greatly surpasses simple listing. If a woman would regard “starting my period” as a favorable event, she would probably make some positive comments about the event, thereby indicating that it was a favorable event. Thus our “judges” were not so blind as Bem (1977) implies.

Bem suggests that the AGT produces instability because of an “open invitation to generate mood or other state-descriptive adjectives.” This is a speculation. Also speculative is the suggestion that “traits that are especially stable might well become ‘ground’ rather than ‘figure.’” If Bem assumes by ground that no traits are likely to appear in subjects’ self-descriptions our 1977 evidence is that traits did frequently appear in subjects’ lists over the 47-day longitudinal period of time. We do not know empirically the exact proportion of generated words that are state words but we
can say that subjects generated more so-called trait words than so-called state words. We also can say that the trait words that subjects used are as apt to be different from one day to the next as are so-called state words. As an example of the proportion of trait versus state words generated in response to the AGT format let us look at the actual adjectives generated by Bem in his rejoinder. Recall that Joe was described as "Consistent, Conservative, Plain Old Average" Joe. And that the woman experiencing the onset of her monthly menstrual period was described as "Feminine and Fulfilled." And that Daryl Bem described himself as "Nurturant, Altruistic, Irreligious, (but not so) Befuddled." Sure, Joe went from the states "Tired" to "Rested" overnight, but only when the AGT was misapplied. That is, AGT instructions routinely ask subjects to generate adjectives "at the end of the day." The reason for this is to enhance the relative constancy of time, place, and situational influences (including natural fatigue); and, to provide subjects with a sufficient framework for evaluating an entire "today," not just the "morning after."

The other discrediting speculation raised by Bem about the AGT is that it "translates relatively trivial changes in an individual's self-description into very large intraindividual variance values." This also is untrue. Bem's example again is hypothetical and misapplies the AGT in its standardized use. A 110-point difference between Nurturant (349) and Altruistic (459) is evaluated as reflecting some "wild bounce." In fact, this difference actually represents a single interval difference in the AGT's 7-point normative rating scale. This difference may be viewed as the AGT's equivalent to Bem's valuing the two adjectives "about equally." The AGT is misapplied in this instance because the difference of 110 points is cited in relation to two adjectives only. The standard number of adjectives contributing to the AGT daily index is five.

Consider the following: Words having the highest and lowest rated favorability on the AGT are Sincere (573) and Liar (26). The 5½-step difference between these two words admittedly would represent a "wild bounce." Note that if a subject were to describe himself at the maximal value five times, the resulting AGT mean would be 573. However, if the same subject were to describe himself only four times at this 573 value and have added the fifth adjective value of 26, the resulting mean then would be
This difference of 110 points does not represent a wild bounce. The difference is just not significant.

Remember Plain Old Average Joe? In fact, he does not "embarrass" (270) us. The wonders of a good Saturday night's sleep did increase the numerical favorability of his self-description but, alas, not significantly so. A simple t-test for correlated means (which assumes some meaning to the constant order of the adjectives "generated" by Joe) indicates that Joe's Saturday-to-Sunday difference of 57 AGT points falls short of being significant \((t = 1.00, df = 4, p > .05)\). "Wild bounce"? Only if one attends to the misleading percentile rankings supplied by Bem does the difference seem large. Note that these percentiles were derived by comparing (not-so-average) Joe's single day means with the 47-day means of real persons. Furthermore, Joe is not like our AGT subjects. Our subjects are complex and interesting. They don't use exactly the same adjectives in the same order with one exception on two consecutive days. Nor would they be likely to consider merely "getting a night's sleep" to be a "significant event."

Instead of the contrived Plain-Old-Average-Joe example we would like to present a real subject. Here are the words a female subject from Allen and Potkay's (1977) study used to describe herself on two consecutive days during the summer of 1973.

<table>
<thead>
<tr>
<th>July 5, 1973</th>
<th>July 6, 1973</th>
</tr>
</thead>
<tbody>
<tr>
<td>lazy</td>
<td>pleased</td>
</tr>
<tr>
<td>indifferent</td>
<td>content</td>
</tr>
<tr>
<td>silly</td>
<td>mature</td>
</tr>
<tr>
<td>unproductive</td>
<td>loved</td>
</tr>
<tr>
<td>apathetic</td>
<td>appreciative</td>
</tr>
<tr>
<td>(M = 174)</td>
<td>(M = 481)</td>
</tr>
</tbody>
</table>

This subject reported that she was ill on July 5, and on July 6 she reported that she talked to her mother on the phone on her mother's birthday. When this same subject had a fruitful day working at home on August 9 her mean AGT was 386. The next day when her period started her AGT mean was 160.

Is it necessarily true that "Allen and Potkay in fact throw away all of this individualized information"? We believe not. Our primary commitment in AGT research has been idiographic. We consistently have sought to permit subjects the opportunity to de-
scribe themselves in their own ways using whatever adjective person-concepts they deem most appropriate. An accompanying experimental commitment of ours has been to minimize unnecessary sources of data confounding. Data contamination and demand character influences would be more likely to occur if subjects were asked not only to generate adjectives but also to rate these same adjectives on an experimenter-designated variable. When subjects rate their own adjectives the experimenter designated variable becomes known to them and in the case where that variable has strong social desirability implications (e.g., favorability) subjects are likely to distort ratings so as to make themselves appear socially desirable. This is why we could not seriously accept Bem's "optimal way to conduct the Allen-Potkay study."

Concerning Bem's "optimal way" suggestion, let us disregard the fact that the two questions posed are different from those we saw as important for our original study. The two questions "How do you feel about yourself today? How good was your day?" increase the likelihood of contamination occurring along a single affective-evaluative experimenter dimension, as contrasted with our own approach. Our value on minimizing confounding also is why we have adopted the practice of using outside raters who are similar to our experimental subjects. Bem, himself, has reported on his own use of such a procedure (1967).

In brief, experimental choices must be made. Allen-Potkay have chosen to give greater priority to subjects generating their own adjective-con structs, and lesser priority to subjects then rating these constructs. Bem-Allen (1974) have chosen to give lesser priority to subjects generating their own trait-constructs (e.g., "friendly and outgoing . . . conscientious"), and greater priority to subjects rating these constructs. Perhaps neither Allen-Potkay nor Bem-Allen have demonstrated a "totally idiographic" approach.

In response to the notion that we have indexed our own a priori construct with the AGT, the following point should be noted. Although the dimension "favorability-unfavorability" has been the dimension most closely identified with the AGT, we recognize that it represents only one possible dimension out of many. We are not solely "interested in" this particular dimension. In fact, we currently are in the process of obtaining normative values for new AGT dimensions, including "anxiety-calm" and
"masculine-feminine," on an expanded list of 2,500 adjectives. One methodological advantage of the AGT over the Bem-Allen methodology is that researchers are permitted to apply new a priori constructs to previously obtained data. Rather than throwing data away, the AGT format will permit researchers to analyze obtained data on more than one dimension. This may occur because subjects in AGT studies contribute their basic adjective responses in an "unidentified" construct framework. From the standpoint of "demand characteristics," it seems highly unlikely to us that subjects would sense our interest in self-descriptive "favorability" and/or "anxiety" and/or "masculine-feminine." And, if they did sense these demands, what adjectives could they generate that would meet the social desirability of all three or more?

First, Bem asserts that a woman who described herself as "feminine" (464) and "fulfilled" (407) upon starting her menstrual period would be marked by the data as "inconsistent," because judges would rate the event "started my period" as unfavorable. Here Bem confuses the "consistency" at issue, intraindividual day-to-day consistency in AGT favorability, with correspondence between AGT favorability and significant event favorability. In terms of the "consistency" at issue, a person is marked by the data as "inconsistent" only when she varies from day-to-day in AGT favorability. The correspondence between AGT favorability and event favorability is another matter.

Second, results of our "demand character" studies were inaccurately described. Results from these studies do not indicate that our subjects were aware of our hypothesis. To find out whether or not our instructions to our subjects both to generate adjectives and to record significant events cued them as to our interest in the relationship between AGT descriptions and events, we provided a new set of subjects with all of the instructions and materials provided our original subjects. This first set of subjects was asked to examine the instructions and materials carefully and then rate the likelihood (100-point likelihood scale) that each of seven hypotheses, including ours (events determine self-descriptions), was the one that we were interested in confirming. Although several of the hypotheses, including our own, were rated as likely to be the one we were interested in confirming, posttests following analysis of variance indicated that there were no significant differences in the likelihood ratings of the several hy-
hypotheses. It was the case that when these same subjects were asked to select a single hypothesis from the seven hypotheses that they thought was the one of interest to us, more subjects than would be expected by chance selected our hypothesis. However, more subjects than would be expected by chance selected another of the seven hypotheses as the one we were interested in confirming. Further, when a second set of subjects was asked to examine the seven hypotheses and pick the one they thought was most valid as an explanation of human behavior without seeing our instructions to our original subjects or knowing anything about our study, many more than would be expected by chance picked our hypothesis. We think that when subjects in the first set picked our hypothesis they were responding to its apparent "validity" rather than to the possible cues in our instructions.

Third, Bem's (1977) analogy between our longitudinal type of study and psychometric "tests" investigating variance across a set of items responded to on a single occasion is entirely inappropriate. As Fiske (1957) indicates, intraindividual variability refers to the differences in responses made on different occasions. Whatever it is that Bem (1977) has discussed is confounded with item variance. Further, the AGT does not involve test items in any way. We have taken great pains to point out the advantages of the AGT over "tests with items" (see Allen, in press; Potkay & Allen, 1973).

Fourth, Bem (1977) makes the point that the Bem and Allen (1974) theory and methods are "almost tautological." They are entirely tautological. Instead of asking their subjects, "Which traits are relevant to you?" as is called for by their theory, they ask subjects, "Are you consistent?" on some trait and if subjects say yes, Bem-Allen assume that those subjects consider the trait relevant. If the subjects say no, Bem-Allen assume that the trait is not relevant to them. Thus, if one were to ask Bem-Allen (1974) "How do you know 'friendliness' is relevant to your 'low variables'?" they must answer, "Because our 'low variables' are consistent." And, if one were to ask Bem-Allen, "Why are your 'low variables' consistent?" they must answer, "Because 'friendliness' is relevant to them." Allen (in press) has noted the meaninglessness of such circularity.

Finally, Bem (1977) committed some logical errors which indicate a lack of understanding of the Allen-Potkay research. Twice Bem (1977; personal communication 3/10/77) has stated
that our results would have been stronger had we "been more idiographic." Of course, he also indicates that our results are not relevant to his theory with Allen (1974) because we were "not idiographic enough." If our method became more idiographic and thus more appropriate to testing his theory, our results would be even stronger evidence against his theory. 

This is because "results being stronger" implies increased correspondence between considerable variance in AGT favorability and considerable variance in event favorability. From Bem’s point of view he should be calling for weaker results the “more idiographic” we become because he believes that becoming more idiographic would attenuate intraindividual variability, the obvious trait position. This mistake essentially is repeated on the last page of his paper (1977). Here he instructs us on how optimally to conduct our kind of study. He comments, “Is there a procedure that . . . will yield a high degree of intraindividual consistency . . . [and is] likely to yield an even higher correlation between the daily measure of ‘favorability toward self’ and the favorability of daily events?” The answer is no, because it is impossible. From his point of view he is correct to call for high intraindividual consistency. Again, it is the trait theorist thing to do. The traits that people have are not supposed to change from day to day according to trait notions. However, for any two variables, as the variance in one is eliminated (high intraindividual consistency) the correlation coefficient approaches zero. The following example illustrates this point.

<table>
<thead>
<tr>
<th>Consecutive days</th>
<th>Daily AGT means for a “subject”</th>
<th>Significant event means</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>451.2</td>
<td>5.5</td>
</tr>
<tr>
<td>2.</td>
<td>451.3</td>
<td>1.2</td>
</tr>
<tr>
<td>3.</td>
<td>451.1</td>
<td>2.5</td>
</tr>
<tr>
<td>4.</td>
<td>451.1</td>
<td>5.6</td>
</tr>
<tr>
<td>5.</td>
<td>451.2</td>
<td>1.4</td>
</tr>
<tr>
<td>6.</td>
<td>451.3</td>
<td>5.8</td>
</tr>
</tbody>
</table>

\[ r = -.08 \]

From his point of view, Bem should have called for a much lower correlation.

We note that Bem (1977) and Bem-Allen (1974) have been
quite willing to rely upon speculation and hypothetical example in their defense of the trait position. "Intuitions or research? One of them must be wrong... we believe that our intuitions are right; the research, wrong" (Bem & Allen, 1974). Our belief is that the research is "right." We are not sure about intuitions. It is incumbent upon Bem-Allen to substantiate their intuitions more with real data and less with absolutisms ("Allen and Potkay obey none of these injunctions.") and hypothetical anecdotes of the "Average Plain Old Joe" variety. Their basic assumption is that people are themselves intuitive trait theorists. Perhaps, it is this basic assumption that is "wrong." Because this assumption is testable, it is possible to determine the extent to which it may or may not be correct. Let us ask people. Let us also ask Bem-Allen to try the AGT—their way. Let us look at the data.

References


Bem, D. J. Predicting more of the people more of the time: Some thoughts on the Allen-Potkay studies of intraindividual variability. *Journal of Personality,* 1977, 45, 327-333.


This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.