Comparing Genuine and Simulated Suicide Notes: A New Perspective

Stephen T. Black

E. S. Shneidman and N. L. Farberow (1957) preselected writers of simulated suicide notes to eliminate vulnerable subjects. Subsequent comparisons of genuine and simulated notes have perpetuated the methodological misstep of the original study. In this study, a new set of genuine notes were selected from completed suicides by men and women who left at least one note, who were White, and who were older than 18 years of age. The simulated note writers (SNWs) were unpreselected, unpaid community volunteers. Genuine note writers in the current and 1957 samples were not found to differ; SNWs from the samples did differ. Problems with the interpretation of differences between genuine and simulated notes are discussed, with a focus on the role-playing nature of the simulated notes.

In the late 1950s, Shneidman and Farberow (1957) hit on the clever notion of comparing genuine and simulated suicide notes for the insight they may provide into the suicides' state and the reasons for their act. The suicide note is, by definition, the product of a person who has decided to die. It is usually written in the minutes or hours immediately preceding the act. Hence, as a personal document, it ought to provide some insight into the individual's psychological state and thoughts. The suicide note is a crucial bit of data that can suggest and illuminate.

Examination of the work in this area reveals a critical methodological problem in the original Shneidman and Farberow (1957) comparison of genuine and simulated suicide notes. For ethical reasons, they preselected simulated-note writers (SNWs) using a personality test and psychological history, eliminating subjects who might have been adversely affected by the role play of writing a simulated suicide note. This confounded the state of being not suicidal with that of psychological stability; any observed differences between the genuine notes and those written by the simulators might have been attributable to differences in history of depression and stability rather than the assumed suicidal versus nonsuicidal state of mind. Because all but one (Darbonne, 1969) of the later comparisons of genuine and simulated notes (Edelman & Renshaw, 1982; Gottschalk & Gleser, 1960; Leenaars & Balance, 1981, 1984; Leenaars, Balance, Wenkstern, & Rudzinsky, 1985; Lester & Leenaars, 1987; McLister & Leenaars, 1988; Ogilvie, Stone, & Schneidman, 1966; Osgood & Walker, 1959; Tripodes, 1976; Tuckman & Ziegler, 1966) were reanalyses of the Shneidman and Farberow (1957) data, the methodological problem of the original study has been perpetuated. The only remedy for this problem is to derive new samples of genuine and simulated suicide notes without the use of preselection.

In this study, I used an unconfounded set of genuine and simulated suicide notes. Primarily I used a content analysis of the manifest content of the text to analyze these notes. For comparison, the Shneidman and Farberow (1957) note pairs were also analyzed using this protocol.

Method

The first sample of genuine and simulated suicide notes used in this study was the original Shneidman and Farberow (1957) set of 33 note pairs. Genuine notes in that sample were selected from 721 suicide notes obtained from the Los Angeles County Coroner's Office, so that all of the note writers were White, native-born, Protestant men between 25 and 59 years of age. The comparison group of simulated notes was created by matching 33 nonsuicidal men with the genuine note writers (GNWs) on these characteristics, and by matching man for man according to age (within 5 years) and occupation. The SNW group was composed of unpaid volunteers solicited from community groups such as local unions and men's fraternal organizations. Volunteers were preselected, before matching, using scores from the Cornell Medical Index. Any potential participant whose score indicated a history of depression, emotional disturbance, or who, in Schneidman and Farberow's (1957) judgments "would be perturbed by having to think of suicide" (p. 253), was asked to write about the happiest experience of his life. Men with acceptable scores on the index were asked to write the "suicide note you would write if you were going to take your own life."

In the 1986 sample of note pairs, genuine notes were obtained from coroner's files in two San Francisco Bay Area counties for calendar years 1985 and 1986. Completed suicides by White adults aged 18 years or older who had left a note in English were included. These criteria resulted in the selection of 77 GNWs: 54 men and 23 women ranging in age from 18 to 90 years.

The SNWs were unpaid volunteers solicited from community groups (e.g., the Junior Chamber of Commerce, the Lions Club, the AFL-CIO Carpenters Union, the Computer Users Group, and community college night school classes) by asking members to participate in a study on suicide. The volunteers were matched person for person with the GNWs according to sex, age group (see Table 1), and occupational level (see Table 1). Occupational level was determined using the Office of Personnel Management's (1983) listing of professional, white-collar,
and blue-collar occupations, with the addition of levels for homemaker and student.

Each participant was interviewed individually. During the interview, he or she was informed that the study involved writing a suicide note and was asked to fill out an informed consent form. Three people declined to participate at this point. No one withdrew from the study after signing the consent form.

Participants were then asked to "write the suicide note you would write if you were going to take your own life." On completing the simulated suicide note, participants were asked to write a second document: "Please write a counselor's response to the person who just wrote that suicide note. Write a response that would prevent the suicide from occurring." No time limit was given for the writing of either document. The purpose of eliciting the counselor's response was to counteract "planting the seed" of suicide ideation, to return participants to their state of mind before the topic of suicide was raised. McGuire and Papageorgis (1961) found that people who refuted a persuasive argument were subsequently strongly resistant to persuasion counteract "planting the seed" of suicide ideation, to return participants to their state of mind before the topic of suicide was raised. McGuire and Papageorgis (1961) found that people who refuted a persuasive argument were subsequently strongly resistant to persuasion

After writing the note and response, participants completed a 12-item suicide information questionnaire about their personal experience with and knowledge of suicide. Immediately after finishing the suicide information questionnaire, participants completed a demographic survey that provided information for matching. Throughout the study, the experimenter was present to monitor participants' reactions to writing the documents and to terminate the session if necessary. Several participants spontaneously said that after writing the counselor's response, they could never contemplate suicide again. Participants did not leave the interview until the experimenter was satisfied that each individual was in a relaxed state. Participants were recontacted 1 week later to check on and relieve any lingering effects of engaging in the study. No one regretted participating or reported feeling any ill effects. Four women said that they had been slightly depressed for a few days afterward.

Both samples of notes were analyzed via a content analysis, which focused on the manifest content of the notes, in terms of thought units rather than words. A thought unit was not defined by punctuation marks but by the expression of a single thought. For example, for the category "affect: depression," the statement "I feel really down in the dumps; things are so sad" would be scored as two occurrences of the category because it connects depression with two different things (i.e., the first, the "I," and the second, "things"). The two phrases form more than one thought. However, for the same category, the statement "I feel really down in the dumps; I'm just so sad" would be scored as a single occurrence of the category because depression is asserted for just one thing, the "I." A single thought unit could be scored as an incidence of more than one category, but it was scored only once for each category.

The content analysis used six major categories: whole-note evaluations; affect; textual content; thought process; values; and residual characteristics. Each major category was divided into subcategories, resulting in a total of 51 content analysis categories. For instance, affect had five subcategories: depression; hostility—intrapersonal; hostility—interpersonal; hostility—impersonal; joy; and love.

Three raters (two undergraduates and I) determined the reliability of the content analysis. All raters were trained using a set of 10 pairs of genuine and simulated notes that were not part of experimental samples. The lowest correlations among the raters were found in the whole-note evaluations, with pairwise correlations in these categories ranging from .66 to .01. Statistical analyses were performed using only variables with a reliability of .64 or greater.

The reliability of scores for the other five categories used in the analysis was satisfactory. The correlation among raters for "residual characteristics: number of thought units" was extremely high, with the lowest pairwise correlations of .96 (between Raters 1 and 2 and between Raters 2 and 3) and the highest of .97 (between Raters 1 and 3). The correlations among raters for the other major categories were also high, ranging from a pairwise high of .91 for affect between Raters 1 and 2 to a low of .64 for values between Raters 2 and 3.

For the actual analysis, I used only the ratings of notes produced by the two suicide theory naive raters. The correlation between these two raters on all categories (except whole note evaluations) was good, ranging from .64 to .96.

Results and Discussion

In the two sampled counties, 34% of all suicides left notes. Comparisons between the GNWs selected for inclusion in the study and all non-note-writing suicide completers in the two counties failed to show significant differences between the two groups in terms of sex, age, marital status, occupational level, employment status, place of death, or method of suicide. Although it cannot be assumed that individuals who leave notes and those who do not are similar in all psychologically relevant respects, no obvious differences (other than the note) were found.

Shneidman and Farberow (1957) found that genuine notes were significantly longer than were simulated notes. The current sample revealed the same difference in length as that shown in the earlier study (the GNWs' mean thought unit was 13.64; the SNWs' mean thought unit was 7.64), F(1, 152) = 11.38, p = .0009. Longer notes provided a greater opportunity for the coded and scored content categories to be used. Spurious differences between the GNWs and SNWs on content categories could therefore result. Because the thought unit and not the word was the primary unit of analysis, I thought it appro-

---

1 The content analysis manual and complete description of its construction are available from Stephen T. Black.
2 All statistical analyses were performed on an Apple SE/30 computer using Systat (Version 3.1, Systat; Evanston, Illinois, 1988) statistical analysis software.
appropriate to control for the length of the note by dividing the number of rated occurrences of each category by the number of thought units in the note.

Comparisons between genuine and simulated notes proceeded in two steps: First, I performed a simple multiple regression analysis, including all content categories as variables, to determine whether there would be a statistically significant relationship between the variables. I then performed univariate $F$ tests to find the specific variables responsible for the observed relationship.

No between-groups or within groups note content differences that were due to sex, age, or occupational level were found in the 1986 sample of GNWs and SNWs. Similarly, no differences that were due to age or occupational level were found in the note content of the Shneidman and Farberow (1957) note pairs.

The comparisons of the 1986 sample of GNWs and SNWs using all content categories as independent variables revealed a highly significant relationship, $F(40, 113) = 3.27$, $R^2 = .537$, $p < .0001$. Univariate $F$ tests showed that SNWs used five content subcategories more often than did the GNWs and that the GNWs used four content subcategories more often than did the SNWs (see Table 2).

If potential confound was manifest in Shneidman and Farberow's (1957) study, the SNWs ought to differ significantly from the 1986 SNWs. However, many other causes, such as cultural change and differences in demographic composition (the 1986 sample included women and had a broader age range than did the original sample), could result in differences. The SNWs and GNWs in the two samples could be compared to eliminate some of the competing hypotheses to explain differences between the samples.

A multiple regression comparison of the GNWs in the Shneidman and Farberow (1957) and the 1986 samples yielded a significant multiple correlation of .649, $F(38, 38) = 1.819$, $p = .034$. Dropping a single variable from the equation—"textual content: relationships"—resulted in the multiple correlation losing significance. This variable seemed likely to have been affected by a cultural change; men of the 1980s might have been more aware of and responsive to interpersonal relationships and, hence, used the category more than did the men of the 1950s.

When the SNWs of the 1957 and 1986 samples were compared, a highly significant multiple correlation of .646 was observed, $F(37, 39) = 3.41$, $p < .0001$; dropping women and multiple note writers resulted in a multiple correlation of .753, $F(37, 39) = 3.21$, $p < .0001$. A univariate analysis of variance showed that the 1957 simulators had higher mean scores on five analysis categories: whole-note evaluations: ambivalence; textual content: asks forgiveness; thought process: use of absolute terms; values: morality of suicide—neutral; and values: morality of suicide—immoral (see Table 2). The single dimension on which the 1986 simulated notes scored higher than 1957 simulated notes was precisely that dimension for which the 1957 writers were rejected: depression.

It seems that the preselection procedure used by Shneidman and Farberow (1957) resulted in a group of SNWs who were good role players. These SNWs were able to produce notes that were highly similar to the genuine notes to which they were compared. The evidence indicates that the preselection procedure confounded the lack of suicidal ideation with an enhanced ability to role play the suicidal state through the selection of adaptable, psychologically healthy individuals as SNWs. This may explain the lack of significant findings that Shneidman (1976) cited as disappointing.

The current sample, which avoided the preselection problem, provided more clearly interpretable information regarding the differences in the content of genuine and simulated suicide notes. The GNWs gave more instructions regarding final affairs, provided more factual information, used more religious ideas, and more often dated the note. GNWs demonstrated a greater awareness of the effect of their death in their concern for giving instructions about final affairs and about information.

### Table 2

<table>
<thead>
<tr>
<th>Content analysis category</th>
<th>$F$</th>
<th>$F$ probability</th>
<th>Dunnnett's $t$ test probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect: depression</td>
<td>5.58*</td>
<td>.01</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Textual content: justification</td>
<td>14.32*</td>
<td>.0002</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Textual content—reasons for suicide: life is overwhelming</td>
<td>12.07*</td>
<td>.0007</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Textual content—reasons for suicide: other</td>
<td>4.57*</td>
<td>.03</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Values—afterlife: none</td>
<td>4.36*</td>
<td>.02</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Textual content—gives instructions: final affairs</td>
<td>8.32*</td>
<td>.004</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Textual content: information</td>
<td>5.49*</td>
<td>.02</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Values: religious ideas</td>
<td>4.51*</td>
<td>.03</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Residual characteristics: date</td>
<td>17.54*</td>
<td>.0001</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

* Simulated note means are greater than genuine note means.  
  b Genuine note means are greater than simulated note means.
tion. The note is the last chance to take care of business, decide who gets what, or to make funeral wishes known. GNWs were aware that they would not be around to tell the reader where the insurance policy is or what bank they use. These things must be related in the note.

Genuine and simulated suicide notes contain differences; the difficulty is in interpreting the differences. Shneidman and Farberow (1957) originally suggested that it is appropriate to compare genuine and simulated suicide notes because by asking non-suicidal people to write suicide notes, a group of notes is created that is equivalent in subject matter and can be further equated by matching note writers on demographic characteristics; because the note sets differ only on the dimension of suicidal–non-suicidal intentions, any observed differences in content between the two sets of notes may be ascribed to the unique psychological state of being suicidal or nonsuicidal.

Shneidman and Farberow's (1957) argument assumes a point that is not in evidence: that any such observed differences in content are important, informative differences. To repeat a cliché, it assumes that they are differences that matter. The nonsuicidal person is asked to role play a suicidal person when he or she writes the fake suicide note. Perhaps the differences reflect only the characteristics that are not easily inferred by the role-playing writer, not necessarily those that are important.

Evaluating the importance of the differences observed in the content of the notes is difficult. Few of the differences have obvious direct clinical relevance. Some of the differences do hang together to provide a sense of the concerns of the suicidal individual, but, in general, the comparison of genuine and simulated notes provides only limited insight into the suicidal state. Clearly, comparing genuine and simulated suicide notes does not put researchers on a royal road to understanding suicide, but such comparisons do highlight the discrepancy between cultural ideas about suicide and the reality of suicide. For example, depression was evident in only 47% of the genuine notes; cultural stereotypes about suicide may lead one to expect a much more frequent occurrence of depression.

Although comparisons of genuine and simulated suicide notes have not been as fruitful as hoped, the study of suicide notes should not be abandoned. Two avenues of investigation may yield more insight into the suicidal state. The contents of genuine suicide notes should be further examined to locate differences associated with demographic characteristics not examined in my study. The comparison of the contents of genuine notes of different age and ethnic groups, and age and sex groups within ethnic groups, may help to explain the variations in the suicide rates of those groups. Attempts to link socioeconomic status to note content have been made (e.g., Shneidman & Farberow, 1960), but such attempts have been frustrated by having available only indirect and aggregate measures of socioeconomic status. The use of occupational level is no substitute for fuller indicators of socioeconomic status such as years of schooling, income, and place of residence.

The second avenue of investigation is to find a comparison group for genuine suicide notes that avoids the problem of role player's intuition. One possibility is to locate the suicide notes of suicide attempters and compare them with the notes of completers. Psychological autopsies may enable the GNWs to serve as their own controls; letters or other personal documents could be compared with the suicide note to illuminate the special state of the suicidal person.

Suicide notes are communications from people who are in a truly suicidal state. In this, the notes are unique sources of information. The study of these notes may yet provide researchers with important insights.

References


