ANAGRAMS: ARS MAGNA?

JUDITH TARR ISQUIT

In her November 1973 Word Ways review of Howard Bergerson's new book, Palindromes and Anagrams (Dover, 1973), Jezebel Q. Xixx declared that "nowhere near 1,169 really good, modern anagrams have been devised". In a letter, she further asserted that she could create a dozen superb anagrams a day indefinitely, all having the quality so conspicuously absent in previously-published examples.

I decided to put this claim to the test. Upon request, Jezebel kindly supplied me with eight anagrams and two antigrams (asterisked) she had recently composed:

- INFERNOS* / non-fires (12.5)
- TANGERINES* / satin-green (19)
- SEXUAL INTERCOURSE / relax, ensure coitus (3)
- ARGUMENTS / must anger (12.5)
- CHRISTIANITY / 'tis in charity (10)
- STREET SHOES / hose testers (18)
- BEHEADMENTS / deathsmen be (16)
- WEIRD NIGHTMARES / withering dreams (3)
- MURMURING WINOS / "rum? rum is now gin!" (20)
- YE SMEARS / are messy (17)

To obtain a comparison set, I searched through the July 1973 to December 1973 issues of the Enigma, the official publication of the National Puzzlers' League, selecting eight anagrams and two antigrams which I felt to be the best of the lot. Mary Youngquist, the present editor of the Enigma, agreed that my choices were reasonable:

- MS. STEINEM / smites men (3)
- GLACIERED / large-iced (10)
- WOMEN'S LIBERATION / men rib as we toil on (7.5)
- PRIMATE / trim ape (6)
- VERSATILITY / variety list (5)
- DIPLOMACY* / mad policy (1)
- PUGILIST / tip: I slug (7.5)
- LEMONADE* / demon ale (10)
- REFURBISH / I rub fresh (14.5)
- ARMAGEDDON / mad god near (14.5)

I shuffled these two lists together, alternating Jezebel's anagrams with the Enigma ones. I then presented the combined list to a sample of 25 people (9 male scientists having Ph.D. degrees, 9 female secretaries, and 7 other people) to evaluate them in view of Jezebel's quality level. To obtain a basis for comparison, each word was scored according to its familiarity with the American population. A score of 100 is equivalent to a word that more than one in a thousand people would recognize on hearing it mentioned. I then gave the score to the next most likely person to recognize it, and so on until the score was assigned. The scores so obtained were the product of the number of people who recognized the word divided by the number of people who were given the word. The highest score is 100.

The results are as follows:

- INFERNOS* / non-fires (12.5)
- TANGERINES* / satin-green (19)
- SEXUAL INTERCOURSE / relax, ensure coitus (3)
- ARGUMENTS / must anger (12.5)
- CHRISTIANITY / 'tis in charity (10)
- STREET SHOES / hose testers (18)
- BEHEADMENTS / deathsmen be (16)
- WEIRD NIGHTMARES / withering dreams (3)
- MURMURING WINOS / "rum? rum is now gin!" (20)
- YE SMEARS / are messy (17)
- MS. STEINEM / smites men (3)
- GLACIERED / large-iced (10)
- WOMEN'S LIBERATION / men rib as we toil on (7.5)
- PRIMATE / trim ape (6)
- VERSATILITY / variety list (5)
- DIPLOMACY* / mad policy (1)
- PUGILIST / tip: I slug (7.5)
- LEMONADE* / demon ale (10)
- REFURBISH / I rub fresh (14.5)
- ARMAGEDDON / mad god near (14.5)

In a few cases, the quality capture rate proved to be low and the score was lower than expected. It is felt that the anagrams are from this sample are possible to be of the same undervalue as the quality of the antigrams are from this sample. However, in rank the score of the male scientists' answers to Jezebel's anagrams does not come in the same undervalue as the quality of the antigrams are from this sample. The population of the Enigma, however, is not "proven" sensitive to their score, and it is clear that it contains a lower quality. Further study and analysis is needed to determine the sensitivity.
Ergerson's Zebel Q. turned anagram that she all having examples. Jezebel (asterisked)

1973 to December the National anagrams which present editor of

It is fairly evident that the low ranks are concentrated in the Enigma anagrams, and the high ranks in the Jezebel ones. Can one conclude from this that the Enigma anagrams are of higher quality? Or, is it possible that both sets of anagrams were drawn at random from the same underlying population of anagram quality, the observed differences in rank being due to the luck of the draw? (By analogy, if one rolls two dice ten times apiece, it is quite likely that the average score of one die will differ somewhat from the average score of the other.) Statisticians have devised a mathematical theory to account for random fluctuations; in particular, they can add up the ranks of the Enigma anagrams and compare this sum with certain calibrated standards. The sum of the Enigma ranks is 79; a number this small or smaller will occur only about 6 per cent of the time if, in fact, the two sets of anagrams come from the same population of anagram quality. Rather than believe that Lady Luck has scored on her first try -- a feat roughly equivalent to rolling either a pair of ones or a pair of sixes with two dice -- the statistician prefers to reject the hypothesis that the two populations of anagram quality are the same, and decides instead that the Enigma population is of higher quality. (Note that statistics has not "proved" this fact in the sense of a mathematical theorem; no matter how small the sum of the Enigma ranks, there is still a possibility that it could have happened with two identical populations of anagram quality.)

Furthermore, it is important to remember that this conclusion applies only to the sample of people used as judges -- or, more precisely, to that population of people of which this sample is representative. Clearly, the United States is not composed of roughly equal parts of male scientists, female secretaries and female Filipinos. These wildly different groups were selected in order to get some idea how sensitive the ranking of the two anagram groups is to the sex and lin-

The results of this survey were fascinating. I scored each anagram +1 if a person put it in the above average group, 0 if in the average group, and -1 if in the below average group. I added up these scores over all 25 people, and ranked the 20 anagrams from 1 (with the highest total score, of +10) to 20 (with the lowest total score, of -16). These ranks are given in parentheses to the right of the anagrams. In a few cases, two (or even three) anagrams had the same total score; these were awarded the average of the ranks involved.

It is fairly evident that the low ranks are concentrated in the Enigma anagrams, and the high ranks in the Jezebel ones. Can one conclude from this that the Enigma anagrams are of higher quality? Or, is it possible that both sets of anagrams were drawn at random from the same underlying population of anagram quality, the observed differences in rank being due to the luck of the draw? (By analogy, if one rolls two dice ten times apiece, it is quite likely that the average score of one die will differ somewhat from the average score of the other.) Statisticians have devised a mathematical theory to account for random fluctuations; in particular, they can add up the ranks of the Enigma anagrams and compare this sum with certain calibrated standards. The sum of the Enigma ranks is 79; a number this small or smaller will occur only about 6 per cent of the time if, in fact, the two sets of anagrams come from the same population of anagram quality. Rather than believe that Lady Luck has scored on her first try -- a feat roughly equivalent to rolling either a pair of ones or a pair of sixes with two dice -- the statistician prefers to reject the hypothesis that the two populations of anagram quality are the same, and decides instead that the Enigma population is of higher quality. (Note that statistics has not "proved" this fact in the sense of a mathematical theorem; no matter how small the sum of the Enigma ranks, there is still a possibility that it could have happened with two identical populations of anagram quality.)

Furthermore, it is important to remember that this conclusion applies only to the sample of people used as judges -- or, more precisely, to that population of people of which this sample is representative. Clearly, the United States is not composed of roughly equal parts of male scientists, female secretaries and female Filipinos. These wildly different groups were selected in order to get some idea how sensitive the ranking of the two anagram groups is to the sex and lin-
guistic background of the judges. It is plausible that males and females view these anagrams through different eyes; two deal with Women's Liberation, one deals with sexual intercourse, and another (street shoes / hose testers) has less meaning for men. If separate rankings are established for male and female judges, differences appear; the sum of the male ranks is 77, and the sum of the female ranks is 89.5. In brief, males see somewhat greater differences between the Jezebel and Enigma anagrams than females do. As might be expected, females view the two Women's Lib anagrams somewhat more positively than males do; however, much larger swings of opinion occur for the sexual intercourse anagram (males +1, females +8), the versatility anagram (males 0, females +7), the arguments anagram (males -3, females +4), the lemonade antigram (males +5, females -3), and the primate anagram (males +6, females -2). I leave to the psychologist the explanation of these differences.

Most of the male-female difference is the contrast in opinion between the native-born males and the native-born females; curiously, the Filipinos by themselves had a rank sum of 79.5, almost the same as the men.

In view of these group differences, it would be foolhardy to conjecture how a random sample of the U.S. population would respond to these anagrams. However, there is, so far, no evidence to support Jezebel's assertions that (1) no high-quality modern anagrams exist, and (2) she can produce high-quality anagrams. If one assumes that the quality of Enigma anagrams in the last half of 1973 is typical of the last 40 years, one can conclude that there exist some 800 published anagrams that (in the eyes of this panel of judges) are at least as good as Jezebel's best.

COMMENT (Jezebel Q. Xixx): A careful and purely objective examination of the anagram evaluation described in the foregoing article provides conclusive proof that the 25 individuals to whom the various anagrams and antigrams were submitted for evaluation were not qualified to pass judgment on them. Just one comparative ranking makes this conclusion indisputably clear.

The INFERNOS / non-fires antigram was assigned a rank of 12.5, while the PRIMATE / trim ape anagram was assigned the much higher rank of 6. No one with a reasonable understanding of anagrams and antigrams could possibly have so ranked these two specimens.

The INFERNOS antigram is one of those rare finds that not merely approaches, but actually reaches, ultimate perfection. An inferno is a fire. To be labeled a non-fire is the summit of anagrammatic (or antigrammatic) success. It leaves nothing to be desired, nothing to be said. There isn't one anagram or antigram in a thousand that exhibits such consummate artistry. By all reasonable standards, it has to rank either 1 or 2 among the 20 examples considered.

The PRIMATE anagram is extremely inferior. First of all, it displays
I am quite certain that no scientist has ever thought of apes as being trim. The most obvious meaning of the word TRIM, when applied to a living being, is in the sense of "slender"; or, possibly, in the sense of "well-manicured". It seems preposterous to think of primates, or of apes, as being well-manicured. As for slenderness, consider the gorilla, a typical ape. The animal is extremely stout, making the descriptive term TRIM totally inappropriate.

Furthermore, this so-called anagram illustrates the common error of confusing an anagram and its base. Apes can be called primates, which all of them are, but primates cannot be called apes, because only some primates are apes. For the anagram even to be considered, however, fleetingly, the construction must be reversed: TRIM APES / primates.

All of these considerations are self-evident to anyone who understands anagrams. The fact that the 25 individuals who judged the two collections of anagrams and antigrams placed the PRIMATE anagram far ahead of the INFERNOS antigram proves, beyond the shadow of a doubt, that their judgment is without any logical foundation. The final results of their ranking bear this contention out.