Psychological Bulletin
1978, Vol. 85, No. 4, 845-857

Gender Effects in Decoding Nonverbal Cues

Judith A. Hall
The Johns Hopkins University

This article summarizes results of 75 studies that reported accuracy for males and females at decoding nonverbal communication. The following attributes of the studies were coded: year, sample size, age of judges, sex of stimulus person(s), age of stimulus person(s), and the medium and channel of communication (e.g., photos of facial expressions, filtered speech). These attributes were examined in relation to three outcome indices: direction of effect, effect size (in *SD* units), and significance level. Results showed that more studies showed female advantage than would occur by chance, the average effect was of moderate magnitude and was significantly larger than zero, and more studies reached a conventional level of significance than would be expected by chance. The gender effect for visual-plus-auditory studies was significantly larger than for visual-only and auditory-only studies. The magnitude of the gender effect did not vary reliably with sample size, age of judges, sex of stimulus person(s), or age of stimulus person(s).

The study of people's ability to judge the meanings of nonverbal cues of emotion has a long history in social psychology, dating to the second decade of this century. The first question asked was whether people could recognize nonverbally expressed emotions at all, and this was followed by the search for correlates of judging ability. Gender was one of the first attributes of judges to be examined in relation to judging ability.

Gender has not always been an important variable in psychological research. In the study of nonverbal communication, however, gender was considered important from the start, because of the predictions that could be made based on gender role stereotypes and folk beliefs about "woman's intuition." Researchers of nonverbal communication have clearly felt that the comparison of males' and females' performances is theoretically interesting and not just a more exhaustive way of reporting the data.

Despite the intuitive and theoretical importance of learning about differences between the sexes in ability to judge nonverbal communication, there exists considerable uncertainty over the facts. Coleman (1949, p. 1) said the existing literature showed "no major sex differences . . . in the identification of emotions," but went on to say that "women tend to be slightly superior to the men in identification of facial expressions." Weisgerber (1956) said the results in the literature were not clear, and Davitz (1964, p. 20) said that "the research evidence about sex differences in sensitivity to facial expression is consistently contradictory." Tagiuri (1969) said the literature was "somewhat contradictory" (p. 406) but that Jenness's (1932b, p. 339) conclusion that "women slightly excel men as judges of facial expressions of emotion" might be "barely right." Tagiuri also suggested (as did Davitz, 1964) that the gender effect may be moderated by the sex of the communicator. In the context of examining judgments of facial expressions among preliterate people, Ekman and Friesen (1971, p. 126) said, "No differences between male and female subjects were expected, and

The author was formerly named Judith Hall Koivumaki.

The author extends many thanks to Robert Rosenthal for his encouragement and advice, and to him, Bert F. Green, Jr., and Stephen M. Kosslyn for criticisms of earlier versions of this article.

Requests for reprints should be sent to Judith A. Hall, Department of Psychology, The Johns Hopkins University, Baltimore, Maryland 21218.

no such differences had been found in the literate culture data." Here, unfortunately, it is unclear precisely what literate culture data were being referred to. Finally, Gitter, Kozel, and Mostofsky (1972) said the results of sex analyses in the literature were mixed.

Maccoby and Jacklin (1974) reviewed sex differences in "empathy" and concluded that, despite the stereotype that girls are more "tuned in" to what other people are thinking and feeling than boys are, there is "no clear tendency" for females to be more sensitive to social cues (p. 211), and "neither sex has greater ability to judge the reactions and intentions of others in any generalized sense" (p. 214). Maccoby and Jacklin's definition of empathy allowed for the inclusion of studies using stimuli other than nonverbal cues and dependent measures other than accuracy at judging. In fact, of the more than 20 studies they reviewed, only two occur in the data set analyzed in the present article. The point is not that Maccoby and Jacklin drew the wrong conclusion[1] but that their conclusion was phrased in language that further clouds our understanding of gender differences in ability to decode nonverbal cues. Further, as has been noted by Block (1976), Maccoby and Jacklin's use of probability level as an index of outcome precludes a full understanding of the directions and magnitudes of the effects.

In a recent article, Hoffman (1977) drew attention to the heterogeneous nature of the studies included by Maccoby and Jacklin under their heading of empathy. In his article, Hoffman made the important distinction between empathy in the sense of vicarious responding and empathy in the sense of recognition of affect. He defined *recognition of affect* as affective perspective taking and under that heading included studies measuring a variety of more cognitive abilities, including verbal report of how a story character feels (e.g., Deutsch, 1975), set to perceive anger in faces (not specifically accuracy at perceiving such cues; Hebda, Peterson, & Miller, 1972), and decoding of nonverbal cues of emotion (e.g., Gitter, Mostofsky, & Quincy, 1971). Thus, his recognition of affect category, like Maccoby and Jacklin's, was heterogeneous and also more inclusive than the category of decoding nonverbal cues that is the subject of the present article. Hoffman, on the basis of his recognition of affect category, concluded that "when encountering someone in an emotional situation, both sexes are equally adept at assessing how that person feels" (p. 716), though several of his studies did not involve direct judging of persons but a general understanding of emotions or social situations, and one did not measure judging accuracy of any kind (Hebda et al., 1972). Finally, Hoffman's conclusion of "no difference" was based on a very incomplete review of the relevant literature, especially in the area of decoding nonverbal cues: Only two of the many published studies on decoding are to be found in Hoffman's review (Gitter et al., 1971; Walton, 1936).

The present article represents an attempt to summarize all the available studies bearing on the question of gender differences in ability to decode nonverbal cues of emotion. The usual interpretative verbal summary is supplanted here, as much as possible, by quantitative analyses of the relationship of direction of effects, effect sizes, and significance levels to attributes of the studies in the data set. The use of effect sizes in summarizing research has recently been demonstrated in reviews by Rosenthal and Rosnow (1975) of attributes of volunteer subjects, by Rosenthal (1976) of outcome of studies on self-fulfilling prophecy, and by Smith and Glass (1977) of psychotherapy outcome studies. In each of these cases, the calculation of effect sizes added appreciably to the author's ability to detect consistencies in the literature.

Method

Description of Data Set

The data set consists of studies that were either published or known to be submitted for publication when this article was written. All the studies measured ability to make correct judgments of emotions or states that were communicated or indicated by nonverbal cues conveyed via face, body, or voice tone. The following were not included: (a) unpublished data, (b) studies that assessed attention or

[1] It is true, however, that Maccoby and Jacklin did not examine all the studies that would have met their definition of empathy.

interest instead of accuracy of judging, (c) studies that assessed accuracy of judging personality, (d) studies in which informative verbal cues were included among the stimuli, (e) studies that did not present any results according to gender, and (f) studies in which a describable and quantifiable criterion of judging accuracy was not available. Describable criteria of accuracy were researcher opinion about the feeling expressed in each stimulus, the communicator's intention as to what feeling was expressed in each stimulus, or modal response (in some studies involving open-ended responses).[2]

In all of the studies in the data set, groups of subjects served as judges of nonverbal expressive stimuli. These stimuli consisted of face, body, or vocal cues (alone or in combination) presented in a variety of media: drawings, photos, films, videotapes, content-standard speech, randomized-spliced speech, and electronically filtered speech. In the method of content-standard speech, the speakers recite meaningless or affectively ambiguous material, varying the expression to suit the intended emotion. In the randomized-spliced method (Scherer, 1971), the audiotape is cut into small pieces and randomly reassembled, creating scrambled speech. Electronic filtering (Rogers, Scherer, & Rosenthal, 1971) removes critical bands of frequencies, making the voice muffled and the words unintelligible.

The expressive stimuli were generated in a variety of ways. By far the most common method was to ask the sender(s) to express various emotions while these expressions were photographed, audiotaped, and so forth. The early studies often employed standard sets of facial-expression stimuli generated in this way. "Spontaneous" expressions were elicited in some studies by presenting subjects with emotion-laden stimuli and recording their unrehearsed responses. For example, Coleman (1949, Studies 21–24 in Table 1 below) used stimuli such as a snake and a crushed snail; Buck (1976, Studies 43–49) and Zuckerman, Hall, DeFrank, and Rosenthal (1976, Study 53) used affectively arousing slides or videotapes. Other stimuli were photos cut out of magazines (e.g., Vinacke, 1949, Study 25), voices of people who were either rested or tired (Fay & Middleton, 1940, Study 57), voices of people who were either lying or telling the truth (Fay & Middleton, 1941, Study 58), and drawings of facial expressions (e.g., Moss, 1929, Study 10).

The stimuli were presented individually and subjects made a judgment immediately following the presentation of each stimulus. Exposure times and interstimulus intervals were often not reported in this literature. Judgments were typically multiple choice, with the response alternatives describing the possible names of the emotion (e.g., happiness, fear, disgust), the interpersonal situation represented in the stimulus (e.g., ordering food in a restaurant, telling the truth or lying), or the nature of the event eliciting the nonverbal expression in the stimulus (e.g., car accident, pleasant adult–child interaction). The number of response alternatives ranged from 2 to over 25. Other judging formats included pointing to a photograph after hearing a story (Ekman & Friesen, 1971, Studies 36 & 37), picking from an array the two faces that expressed the most similar emotions (Fields, 1953, Study 26), open-ended responses (Gates, 1923, Study 1), and rating degrees of friendliness (Zaidel & Mehrabian, 1969, Study 69).

In all but one of these studies, gender effects were reported or commented on. The exception was the study by Zuckerman and Przewuzman (Note 1, Study 54); in that study, the data-analytic model made it clear that gender effects were examined, and since no significant effect was reported, it was inferred that no significant effect was obtained.

Many studies that were examined did not report or comment on gender effects. There is no way of knowing if such effects were never looked at or if they were looked at, found to be nonsignificant, and then forgotten. If the latter is the case, then it is conceivable that the data set is biased in having a disproportionate number of studies showing large and/or significant effects. Whether there might be a bias to overrepresent significant effects favoring one sex over the other is a different question. Such a bias seems unlikely, but if present would, of course, create a greater problem for inference in this article than would a bias to include significant effects ignoring direction.

Each entry in the data set represents an independent sample or group of samples, except for Studies 16 and 56 (Dusenbury & Knower, 1938, 1939), Studies 50 and 64 (Rosenthal, Hall, DiMatteo, Rogers, & Archer, in press), and Studies 65 and 66 (Rosenthal et al., in press). In these cases, the same samples were given two different tests and were entered separately in the data set in order not to sacrifice results for the different tests. In addition, in the book by Rosenthal et al. (in press), gender effects were reported for many independent samples. In these cases, the reported results of all samples of the same age group that were administered the same test were pooled and not entered as individual samples, as is explained in Footnotes f–i in Table 1.

Attributes Recorded

For all of the studies in the data set, the following attributes were recorded and later coded into the categories explained in Table 1: year, sample size, age

[2] One study (Walton, 1936) involved four tests, only one of which measured decoding of nonverbal cues; the other three involved understanding the affective connotations of lines and colors. To this author's reading, Walton's statement of "no sex differences" was made in the context of one of these other tests, though he was very imprecise about this and there is room for disagreement about what he meant. Hoffman (1977) included the study in his review of "recognition of affect," citing the "no difference" result, but it is not included in the present review for the reason stated above.

Table 1
List of Studies, Their Attributes, and Results

Study number	Study	*N*	Age of subjects	Sex of sender(s)	Age of sender(s)	Channel	Direction	Effect (in *SD*)
			Visual mode					
1	Gates (1923)	458	1[a]	2	2	2	1	—
2	Allport (1924)	—	4	1	2	2	1	—
3	Buzby (1924)	207	4	1	2	1	3	—
4	Buzby (1924)	501	4	1	2	1	3	—
5	Fernberger (1927)	770	4	1	2	1	1	—
6	Fernberger (1927)	743	4	1	2	1	1	—
7	Fernberger (1927)	724	4	1	2	1	1	—
8	Hunt (1928)	430	4	3	3	1	3	—
9	Guilford (1929)	15	4	1	2	2	1	.00
10	Moss (1929)	1,000	4	3	3	1	3	—
11	Kanner (1931)	409	4	2	2	2	3	.14
12	Kellogg & Eagleson (1931)	332	1[a]	2	2	2	3	.53
13	Jenness (1932a)	536	4	1	2	2	3	.48*
14	Carmichael, Roberts, & Wessell (1937)	348	4	1	2	4	3	.02
15	Carmichael, Roberts, & Wessell (1937)	292	4	1	2	4	3	.05
16	Dusenbury & Knower (1938)	388	4	3	2	3	3	1.02*
17	Dickey & Knower (1941)	233	2	3	2	2	3	—
18	Dickey & Knower (1941)	70	2	3	2	2	3	—
19	Dickey & Knower (1941)	389	3	3	2	2	3	—
20	Dickey & Knower (1941)	238	3	3	2	2	3	—
21	Coleman (1949)	133	4	2	2	3	3	.10
22	Coleman (1949)	77	4	2	2	3	2	−.31
23	Coleman (1949)	69	4	1	2	3	2	—
24	Coleman (1949)	100	4	1	2	3	2	−.12
25	Vinacke (1949)	600[b]	4	3	2[b]	2	3	.14
26	Fields (1953)	200	4	3	2	2	2	−.15
27	Frijda (1953)	40	4	2	2	8	1	.00
28	Vinacke & Fong (1955)	330[b]	4	3	3	2	2	−.02
29	Weisgerber (1956)	95	4	2	2	2	3	.67*
30	Weisgerber (1956)	100	4	1	2	4	3	.40*
31	N. Levy & Schlosberg (1960)	390	4	2	2	2	3	.11
32	Staffieri & Bassett (1970)	60	1	1	2	2[c]	3	.40*
33	Staffieri & Bassett (1970)	49	1	2	1	2[c]	3	.69*
34	Staffieri & Bassett (1970)	60	1	2	2	2[c]	2	−.60*
35	Staffieri & Bassett (1970)	59	1	1	1	2[c]	1	—
36	Ekman & Friesen (1971)	130	1[d]	3	3	2	3	—
37	Ekman & Friesen (1971)	189	4	3	3	2	3	—
38	Gitter, Mostofsky, & Quincy (1971)	80	1	2	2	2	1	—
39	Izard (1971)	268	4	3	2	2	3	.30*
40	Gitter, Black, & Mostofsky (1972a)	48	4	3	2	2	3	—

Table 1 (*continued*)

Study number	Study	N	Age of subjects	Sex of sender(s)	Age of sender(s)	Channel	Direction	Effect (in SD)
	Visual mode (*continued*)							
41	Gitter, Black, & Mostofsky (1972b)	160	4	3	2	2	3	.26
42	Buck, Miller, & Caul (1974)	64	4	3	2	3	3	—
43	Buck (1976)	60	4	3	2	3	3	.54*
44	Buck (1976)	10	3	3	2	3	1	—
45	Buck (1976)	7	4	3	2	3	1	—
46	Buck (1976)	23	4	3	2	3	2	—
47	Buck (1976)	53	4	3	2	3	1	—
48	Buck (1976)	10	4	3	2	3	3	1.86*
49	Buck (1976)	10	4	3	2	3	3	—
50	Rosenthal, Hall, DiMatteo, Rogers, & Archer (in press)	275	1	2	2	3	3	.03
51	Rosenthal et al. (in press)	92	4	2	2	3	3	.30
52	Sweeney & Cottle (1976)	100	4	3	2	9	3	1.31*
53	Zuckerman, Hall, DeFrank, & Rosenthal (1976)	60	4	3	2	3	3	.65*
54	Zuckerman & Przewuzman (Note 1)	77	1	3	2	2	1	—
55	Zuckerman & Przewuzman (Note 1)	92	4	3	1	2	3	.60
	Auditory mode							
56	Dusenbury & Knower (1939)	388	4	3	2	5	3	.34*
57	Fay & Middleton (1940)	46	4	3	2	5	2	−.17
58	Fay & Middleton (1941)	47	4	3	2	5[e]	3	—
59	Pfaff (1954)	201	4	1	2	5	3	—*
60	Pfaff (1954)	103	2	1	2	5	1	—
61	Beldoch (1964)	89	4	3	2	5	2	−.21
62	Dimitrovsky (1964)	224	1	3	2	5	3	.38*
63	P. K. Levy (1964)	74	4	3	2	5	2	−.14
64	Rosenthal et al. (in press)	275	1	2	1	6	3	.31*
65	Rosenthal et al. (in press)	119	3	2	2	6	3	.03
66	Rosenthal et al. (in press)	119	3	1	2	6	3	.29
67	Rosenthal et al. (in press)	80	4	1	2	6	3	.67*
68	Rosenthal et al. (in press)	131	4	1	2	6	3	.32
	Visual-plus-auditory mode							
69	Zaidel & Mehrabian (1969)	72	4	3	2	7	3	3.28*
70	Gitter, Kozel, & Mostofsky (1972)	183	4	2	2	7	3	.46*
71	Zuckerman, Lipets, Koivumaki, & Rosenthal (1975)	30	4	3	2	7	3	1.12*

(*table continued*)

Table 1 (*continued*)

Study number	Study	N	Age of subjects	Sex of sender(s)	Age of sender(s)	Channel	Direction	Effect (in SD)
			Visual-plus-auditory mode (*continued*)					
72	Rosenthal et al. (in press)	200	1	2	2	7	3	.78[f]*
73	Rosenthal et al. (in press)	581	3	2	2	7	3	.56[g]*
74	Rosenthal et al. (in press)	1,725	4	2	2	7	3	.36[h]*
75	Rosenthal et al. (in press)	237	2	2	2	7	3	.61[i]*

Note. The coding in the table columns is as follows: For Age of subjects, 1 = preschool and grade school, 2 = junior high school, 3 = high school, 4 = college and older; for Sex of sender(s), 1 = male, 2 = female, 3 = both male and female; for Age of sender(s), 1 = children, 2 = college and older, 3 = both children and college and older; for Channel, 1 = drawings of faces, 2 = photos of faces, 3 = moving face or face and body (film or videotape), 4 = photos or film of hands and arms, 5 = standard content speech, 6 = electronically filtered and randomized-spliced (scrambled) speech, 7 = both 5 or 6 and 1, 2, 3, or 4, 8 = both 2 and 3, 9 = photos of faces, postures, hands and arms, legs and feet; for Direction, 1 = no difference, 2 = male advantage, 3 = female advantage; for Effect, positive values indicate female advantage, negative values indicate male advantage.
[a] Ages 3–14 years.
[b] Best estimate.
[c] "Pictures."
[d] Children whose ages were unspecified.
[e] Vocal responses indicating truth telling or lying.
[f] *Mdn* of 10 samples; N is the sum of all 10 samples; test of significance performed with $N-2$ *df*.
[g] *Mdn* of 12 samples; N is the sum of all 12 samples; test of significance performed with $N-2$ *df*.
[h] *Mdn* of 34 samples; N is the sum of all 34 samples; test of significance performed with $N-2$ *df*.
[i] *Mdn* of 2 samples; N is the sum of the 2 samples; test of significance performed with $N-2$ *df*.
* $p \leq .05$, two-tailed.

group of judges, sex of stimulus person(s), age group of stimulus person(s), and the channel of communication and medium (e.g., facial expressions shown on videotape). The following outcomes were recorded and coded: direction of the gender effect,[3] magnitude of the effect (if reported or calculable), and significance level (if known).

Definition and Computation of Effect Size

Effect size tells how large an effect is, disregarding its significance level and its N. The effect size estimate used here is d. It is a measure, in standard deviation units, of the degree of departure from the null hypothesis and is defined as the difference between the means of the two groups divided by their common standard deviation (Cohen, 1969).

A few studies reported effect sizes in terms of d (SD units). In others, it was possible to compute d, using formulas given by Friedman (1968) and Cohen (1969), if any of the following were provided: F, t, U, critical ratio, point-biserial r, or z; means, Ns, and standard deviations for males and females; or proportions of males and females who correctly judged each stimulus. In each case the effect size estimate was signed after computation to show the direction of the effect, with positive values indicating female advantage and negative values indicating male advantage.

Since d is affected by the overall amount of variability in the data, in poorly controlled or less precise studies the signal (mean difference) may be undetectable because of too much noise (variability). Thus d may tell us as much about the nature of the study itself as about the true magnitude of the effect under examination, since it is affected both by variation between groups and by variation among members. Trends in the obtained magnitude of effects over time thus might be attributable either to a reduction in variability due to methodology (reduced measurement error, sampling from more homogeneous populations) or to an actual change in the magnitude of the effect due possibly to some sociological or educational changes that could differentially affect the groups' performance on such tasks.

[3] As Table 1 shows, for two studies, the phrase "no difference" was literally true ($d = .00$ SD). In all other studies showing "no difference," the author merely stated there was no difference or little difference without providing information on direction or magnitude.

Table 2
Summary of Attributes of Studies

Attribute	Modality: Visual (N = 55)	Auditory (N = 13)	Visual plus auditory (N = 7)	All (N = 75)
Median year	1955	1964	1976	1964
Median N	115	119	200	124
Age of subjects[a]	10/2/3/40 (18/4/5/73)	2/1/2/8 (15/8/15/62)	1/1/1/4 (14/14/14/57)	13/4/6/52 (17/5/8/69)
Sex of sender(s)[a]	15/13/27 (27/24/49)	5/2/6 (38/15/46)	—/5/2 (—/71/28)	20/20/35 (27/27/47)
Age of sender(s)[a]	3/47/5 (5/85/9)	1/12/— (8/92/—)	—/7/— (—/100/—)	4/66/5 (5/88/7)

[a] Entries are counts in each of the coding categories described in Table 1. Percentages are given in parentheses.

Results

Table 1 lists all the studies in the data set, with attributes and outcomes numerically coded. Codes for the numeric categories are given in the note to Table 1. Table 2 provides frequencies for each category of each coded attribute, separately for each of the three communication modalities (visual, auditory, and visual plus auditory). Studies using only auditory cues were conducted more recently than studies using only visual cues, and studies involving both modalities tended to be done much more recently. For all three modalities together, studies were most frequently performed with college-age and older subjects (mainly college age) and used stimuli consisting of college-age or older stimulus persons (mainly college age). More studies used both male and female stimulus persons than either sex alone.

Table 3 shows the proportion of studies in which females were better judges of nonverbal cues than males were. These proportions are defined in two ways: (a) including and (b) excluding studies showing "no difference." These proportions are given both for all studies and for only those studies for which the effect size was known. It can be seen that the inclusion of the 14 studies showing "no difference" when only the direction of the gender effect was known resulted in the lowest proportion of studies showing female advantage

Table 3
Proportion of Studies Showing Female Advantage

Definition	Direction of effect known: %	Proportion	Size of effect known: %	Proportion
Female advantage vs. male advantage and no difference	68	51/75[a]	78	36/46[b]
Female advantage vs. male advantage, excluding no difference	84	51/61[c]	82	36/44[d]

[a] $\chi^2(1) = 9.01, p < .01$ (expected values set at .50 of N).
[b] $\chi^2(1) = 13.59, p < .001$ (expected values set at .50 of N).
[c] $\chi^2(1) = 26.23, p < .001$ (expected values set at .50 of N).
[d] $\chi^2(1) = 16.57, p < .001$ (expected values set at .50 of N).

Table 4
Magnitude of Effect (in SD *units) Favoring Females*

	Modality			
Measure	Visual ($N = 29$)	Auditory ($N = 10$)	Visual plus auditory ($N = 7$)	Total ($N = 46$)
M (unweighted)	.32	.18	1.02	.40[a]
95% confidence interval (*SD* units)	.14–.51	−.02–.39	.08–1.97	.22–.58
Mdn	.26	.30	.61	.32
N of subjects	5,671	1,545	3,028	10,244

[a] Weighted $M = .34$ *SD*.

(68%). These 14 studies reporting "no difference" probably did in fact show one sex to exceed the other in accuracy; that is, in most cases the difference was probably not exactly zero. Based on the other proportions shown in Table 3, the figure of 68% is probably quite an underestimate of the true proportion. Regardless of the means of figuring the proportion, Table 3 reveals that the proportion of studies showing female advantage significantly exceeded chance.

In Table 4, effect sizes are shown for the 46 studies for which an effect size estimate was available or could be computed. Below the mean effect size, the table shows the 95% confidence interval for the mean gender effect, in *SD* units. These confidence intervals generally put the population mean above .00, indicating that the true effect is in favor of females. The average effect size (.40 *SD*) is of moderate magnitude (Cohen, 1969) and is very significantly greater than .00, $t(45) = 4.38$, $p < .0001$. The effect size for the visual-plus-auditory category is significantly greater than the mean of the two other modes, $t(43) = 3.23$, $p = .002$. These studies tended to be very recent and probably employed more precise measuring instruments than did earlier studies. The number of stimuli used in these visual-plus-auditory studies tended to be large (often over 100), and the greater reliability of such sets of stimuli would allow for larger effects to be detected. In addition, it can be noted that the visual-plus-auditory stimuli probably had more ecological validity than did the visual or auditory stimuli alone. The greater realism of these stimuli may have resulted in effect sizes that are more representative of gender effects in everyday judgments of nonverbal cues than are the effects obtained for the single modalities.

Table 5
Numbers of Studies Reaching Conventional Level of Significance in Favor of Females ($p \leq .05$, *two-tailed)*

Definition	%	Proportion	χ^2
All studies	31	23/75	65.62[a]*
Studies for which effect size was known	48	22/46	61.09[b]*
Statistically significant studies	96	23/24	18.38[c]*

[a] Based on expected values of 5 and 70, or .07 and .93 of *N*, to meet minimum cell-size requirement of chi-square test; expected values would otherwise have been .025 and .975 of *N*.
[b] Based on expected values of 5 and 41, or .11 and .89 of *N*; expected values would otherwise have been .025 and .975 of *N*.
[c] Expected values set at .50 of *N*.
* $p < .001$.

Turning to significance level, we see in Table 5 that the proportion of studies reaching significance ($p \leq .05$) in favor of females far exceeded chance in three alternative analyses. When significance level was examined separately for the three communication modalities, the percentages (proportions) reaching significance in favor of females were as follows: visual—20% (11/55), auditory—38% (5/13), visual plus auditory—100% (7/7). For all 75 studies, the proportion showing significant female advantage was

Table 6
Correlations of Results of Studies with Study Attributes

Attribute of study	Result of study: Size of effect	Direction of outcome (Index 1)	Direction of outcome (Index 2)	Significance of effect
Year	.28 (44)*	.05 (59)	.15 (73)	.34 (73)**
N	−.07 (44)	.21 (59)	.12 (72)	.04 (72)
Sex of sender(s)	.02 (26)	.05 (28)	.21 (38)	.21 (38)
Age of sender(s)	−.05 (43)	−.10 (54)	−.04 (68)	−.08 (68)
Modality	−.14 (37)	−.09 (52)	.05 (66)	.15 (66)
Age of subjects	.06 (44)	−.13 (59)	.02 (73)	−.10 (63)

Note. The coding in the table is as follows: For Size of effect, higher values indicate greater female advantage; for Direction of outcome (Index 1), 1 = male advantage, 2 = female advantage; for Direction of outcome (Index 2), 1 = male advantage and no difference, 2 = female advantage; for Significance of effect, 1 = $p > .05$, 2 = $p \leq .05$; for Sex of sender(s), 1 = male sender(s) only, 2 = female sender(s) only; for Age of sender(s), 1 = child(ren), 2 = adult(s); for Modality, 1 = visual modality, 2 = auditory modality; for Age of subjects, 1 = preschool and grade school, 2 = junior high school, 3 = high school, 4 = college and older. Degrees of freedom for the correlations are in parentheses.
* $p < .10$.
** $p < .005$.

12 times what would be expected by chance, and the combined probability was .00003, two-tailed.

As suggested above, the 75 studies may overrepresent significant effects if studies showing such effects were more likely to be published than studies yielding nonsignificant effects. A quick calculation reveals that over 300 more studies showing nonsignificant gender effects ($p > .05$) would have to exist somewhere in order for χ^2 to fall to .00. This seems unlikely. Nevertheless, there may still be some bias toward overrepresentation of significant effects in the 75 studies reviewed here. However, such a bias cannot explain why so many of the published studies have shown females to outperform males.

Table 6 displays correlations of study attributes (dichotomously coded in some instances, as explained in the table's notes) with results of studies. What is most apparent about this table is the generally small magnitude of the correlations, which attests to the robustness of the female advantage. More recent studies tended to show larger effects (and to be more significant, despite a negative r between year and sample size, $r(72) = -.38$, $p = .001$). However, the magnitude of the gender effect was not reliably predicted by sample size, sex of sender(s), age of sender(s), communication modality (visual vs. auditory), or age of subjects. The lack of relationship between effect size and age of subjects ($r = .06$) is supported by the failure of internal analyses reported in some of the studies to detect appreciable interactions of sex of subject and age of subject (Dimitrovsky, 1964, Study 62; Rosenthal et al., in press, Studies 72–75).

The slight tendency shown in Table 6 for the gender effect to be larger for visual than auditory stimuli ($r = -.14$) occurred much more powerfully in the study by Zuckerman, Lipets, Koivumaki, and Rosenthal (1975, Study 71), and in the one by Rosenthal et al. (in press, Study 73), where significant interactions of modality (visual versus auditory) and sex of subjects were reported, in each case showing the gender effect to be greater for visual stimuli than for auditory stimuli.

The correlations in Table 6 showing that sex of sender made little difference are consistent with internal analyses reporting no interactions of sender's and judge's sex in some of the studies in the data set (Dimitrovsky, 1964, Study 62; Dusenbury & Knower, 1938, Study 16; Ekman & Friesen, 1971, Study 37; Gitter, Black, & Mostofsky, 1972a, 1972b, Studies 40 & 41; Rosenthal et al., in press, Studies 65–68; Vinacke & Fong, 1955, Study 28).

Discussion

The analyses revealed that more studies showed female advantage than would occur by chance, the average effect size was significantly larger than .00, and more studies reached the $p \leq .05$ level of significance in favor of females than would be expected by chance. The magnitude of the gender effect has increased with time, most likely due to a combination of more precise measuring instruments and more powerful data analysis. The effect size was appreciably greater for visual-plus-auditory studies than for the single modalities. The magnitude of the gender effect did not vary reliably with sample size, sex of sender(s), age of sender(s), or age of judges. That the gender effect did not vary appreciably with the sex of the sender is inconsistent with hypotheses put forth by Davitz (1964) and Tagiuri (1969).

The most theoretically interesting finding was the absence of an interaction of sex of judge and age of judge, as shown by the correlation of .06 in Table 6 between effect size and age of subjects. If the effect grew either smaller or larger with age, we would be guided toward one or another set of hypotheses regarding the origins of this gender difference. As it is, we can only say that cross-sectionally, the female advantage at judging nonverbal cues is stable.

The direction of the effect is consistent with gender role stereotypes (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972), and it is possible that females learn early (very early, judging from the results reported here) "how a girl ought to act." This learning would probably not directly produce a performance advantage in judging nonverbal cues, but over time the added motivation to relate to others expressively and practice at attending to interpersonal expression might result in females' superior judging ability. It is also possible that nonverbal sensitivity is especially socially adaptive for females; English (Note 2) and Weitz (1974) have both suggested that the oppressed status of women might make sensitivity to nonverbal communication particularly important if it allows them to read better the wishes of more powerful others. A person who has less than optimal social power may learn to act on and employ subtle cues in order to effect more social control. However, that young girls should be better judges of nonverbal communication than young boys are is inconsistent with this hypothesis, unless one seriously believes that very young girls as well as women are oppressed in our society. If either of the social-learning hypotheses is true—the gender role socialization hypothesis or the dealing-with-oppression hypothesis—then the gender effect should diminish if our society is able gradually to shed its ingrained gender role stereotypes and patterns of male dominance. This review should therefore be updated from time to time.

Another kind of explanation, in its simplest form, would hold that females are "wired" from birth to be especially sensitive to nonverbal cues or to be especially quick learners of such cues. This would make evolutionary sense, because nonverbal sensitivity on a mother's part might enable her to detect distress in her youngsters or threatening signals from other adults, thus enhancing the survival chances of her offspring. Relevant to this idea, Rosenthal et al. (in press) reported that females in grade school, high school, and college samples were especially superior at judging negative affective cues. It would be interesting, though methodologically difficult, to learn whether other primate females might also show superior ability to judge nonverbal cues and to see if their special sensitivity is also for angry or sad cues.

The present review clarifies a previously confused issue and establishes that on a variety of conceptually similar tasks of decoding nonverbal cues females are reliably more accurate than males. The mean difference of .40 *SD* is not, however, a large difference. An average effect of this size means that the upper 50% of the female distribution exceeds about 65% of the male distribution, or that about 27% of the combined area covered by the two populations is not overlapped (Cohen, 1969). Less than 4% of the variance in decoding scores is accounted for by gender. Even if only the 23 studies conducted since 1970 are considered, the average effect of .52 *SD* is equivalent to an r^2 of only about .06. However, an effect of this size is said by Cohen (1969) to be "large enough to be visible to the naked eye" (p. 24).

Questions for further study are whether the difference is ameliorable, and at what age the difference emerges. It is also important to keep in mind that every gender difference is a problem in construct validity. Because we do not experimentally manipulate gender, we have to deal with many variables that may be correlated (confounded) with gender, such as personality, socialization, and physical characteristics. Along these lines, research is currently underway that examines whether sex-role-related personality differences may account for the gender difference in decoding nonverbal cues.

Recent research has indicated that sensitivity to nonverbal cues is positively associated with, among other things, interpersonal adequacy, maturity, and achievement potential as measured by the California Psychological Inventory, teaching skill as measured by supervisors' ratings, and clinical skill as measured by peer and supervisor ratings and patient outcomes (Burruss, 1977; Rosenthal et al., in press). Since nonverbal sensitivity thus appears to be an asset, perhaps our educational programs should include training in nonverbal as well as verbal communication skills. The results of the present review would suggest that in such a program, special attention should be given to improving boys' nonverbal decoding skills.

Reference Notes

1. Zuckerman, M., & Przewuzman, S. J. *Decoding and encoding facial expressions in pre-school age children.* Unpublished manuscript, University of Rochester, 1976.
2. English, P. W. *Behavioral concomitants of dependent and subservient roles.* Unpublished manuscript, Harvard University, 1972.

References

Allport, F. H. *Social psychology.* Boston: Houghton Mifflin, 1924.

Beldoch, M. Sensitivity to expression of emotional meaning in three modes of communication. In J. R. Davitz (Ed.), *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Block, J. H. Issues, problems, and pitfalls in assessing sex differences: A critical review of *The psychology of sex differences. Merrill-Palmer Quarterly,* 1976, *22,* 283–308.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. Sex-role stereotypes: A current appraisal. *Journal of Social Issues,* 1972, *28,* 59–78.

Buck, R. A test of nonverbal receiving ability: Preliminary studies. *Human Communication Research,* 1976, *2,* 162–171.

Buck, R., Miller, R. E., & Caul, W. F. Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology,* 1974, *30,* 587–596.

Burruss, J. A. *Evaluating therapists' effectiveness.* Unpublished doctoral dissertation, Harvard University, 1977.

Buzby, D. E. The interpretation of facial expression. *American Journal of Psychology,* 1924, *35,* 602–604.

Carmichael, L., Roberts, S. O., & Wessell, N. Y. A study of the judgment of manual expression as presented in still and motion pictures. *Journal of Social Psychology,* 1937, *8,* 115–142.

Cohen, J. *Statistical power analysis for the behavioral sciences.* New York: Academic Press, 1969.

Coleman, J. C. Facial expressions of emotion. *Psychological Monographs,* 1949, *63*(1, Whole No. 296).

Davitz, J. R. A review of research concerned with facial and vocal expressions of emotion. In J. R. Davitz (Ed.), *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Deutsch, F. Effects of sex of subject and story character on preschoolers' perceptions of affective responses and intrapersonal behavior in story sequences. *Developmental Psychology,* 1975, *11,* 112–113.

Dickey, E. C., & Knower, F. H. A note on some ethnological differences in recognition of simulated expressions of the emotions. *American Journal of Sociology,* 1941, *47,* 190–193.

Dimitrovsky, L. The ability to identify the emotional meaning of vocal expressions at successive age levels. In J. R. Davitz (Ed.), *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Dusenbury, D., & Knower, F. H. Experimental studies of the symbolism of action and voice: I. A study of the specificity of meaning in facial expression. *Quarterly Journal of Speech,* 1938, *24,* 424–435.

Dusenbury, D., & Knower, F. H. Experimental studies of the symbolism of action and voice: II.

A study of the specificity of meaning in abstract tonal symbols. *Quarterly Journal of Speech,* 1939, *25,* 67–75.

Ekman, P., & Friesen, W. V. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology,* 1971, *17,* 124–129.

Fay, P. J., & Middleton, W. C. The ability to judge the rested or tired condition of a speaker from his voice as transmitted over a public address system. *Journal of Applied Psychology,* 1940, *24,* 645–650.

Fay, P. J., & Middleton, W. C. The ability to judge truth-telling, or lying, from the voice as transmitted over a public address system. *Journal of General Psychology,* 1941, *24,* 211–215.

Fernberger, S. W. Six more Piderit faces. *American Journal of Psychology,* 1927, *39,* 162–166.

Fields, S. J. Discrimination of facial expression and its relation to personal adjustment. *Journal of Social Psychology,* 1953, *38,* 63–71.

Friedman, H. Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin,* 1968, *70,* 245–251.

Frijda, N. H. The understanding of facial expression of emotion. *Acta Psychologica,* 1953, *9,* 294–362.

Gates, G. S. An experimental study of the growth of social perception. *Journal of Educational Psychology,* 1923, *14,* 449–461.

Gitter, A. G., Black, H., & Mostofsky, D. Race and sex in the communication of emotion. *Journal of Social Psychology,* 1972, *88,* 273–276. (a)

Gitter, A. G., Black, H., & Mostofsky, D. Race and sex in the perception of emotion. *Journal of Social Issues,* 1972, *28*(4), 63–78. (b)

Gitter, A. G., Kozel, N. J., & Mostofsky, D. I. Perception of emotion: The role of race, sex, and presentation mode. *Journal of Social Psychology,* 1972, *88,* 213–222.

Gitter, A. G., Mostofsky, D. I., & Quincy, A. J., Jr. Race and sex differences in the child's perception of emotion. *Child Development,* 1971, *42,* 2071–2075.

Guilford, J. P. An experiment in learning to read facial expression. *Journal of Abnormal and Social Psychology,* 1929, *24,* 191–202.

Hebda, M. E., Peterson, R. A., & Miller, L. K. Aggression anxiety, perception of aggressive cues, and expected retaliation. *Developmental Psychology,* 1972, *7,* 85.

Hoffman, M. L. Sex differences in empathy and related behaviors. *Psychological Bulletin,* 1977, *84,* 712–722.

Hunt, T. The measurement of social intelligence. *Journal of Applied Psychology,* 1928, *12,* 317–334.

Izard, C. E. *The face of emotion.* New York: Appleton-Century-Crofts, 1971.

Jenness, A. Differences in the recognition of facial expression of emotion. *Journal of General Psychology,* 1932, *7,* 192–196. (a)

Jenness, A. The recognition of facial expressions of emotion. *Psychological Bulletin,* 1932, *29,* 324–350. (b)

Kanner, L. Judging emotions from facial expressions. *Psychological Monographs,* 1931, *41*(3, Whole No. 186).

Kellogg, W. N., & Eagleson, B. M. The growth of social perception in different racial groups. *Journal of Educational Psychology,* 1931, *22,* 367–375.

Levy, N., & Schlosberg, H. Woodworth Scale values of the Lightfoot pictures of facial expression. *Journal of Experimental Psychology,* 1960, *60,* 121–125.

Levy, P. K. The ability to express and perceive vocal communications of feeling. In J. R. Davitz (Ed.), *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences.* Stanford, Calif.: Stanford University Press, 1974.

Moss, F. A. *Your mind in action: Applications of psychology.* Boston: Houghton Mifflin, 1929.

Pfaff, P. L. An experimental study of the communication of feeling without contextual material. *Speech Monographs,* 1954, *21,* 155–156. (Abstract)

Rogers, P. L., Scherer, K. R., & Rosenthal, R. Content filtering human speech: A simple electronic system. *Behavior Research Methods and Instrumentation,* 1971, *3,* 16–18.

Rosenthal, R. *Experimenter effects in behavioral research* (Enlarged ed.). New York: Irvington, 1976.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. *Sensitivity to nonverbal communication: The PONS Test.* Baltimore, Md.: Johns Hopkins University Press, in press.

Rosenthal, R., & Rosnow, R. L. *The volunteer subject.* New York: Wiley, 1975.

Scherer, K. R. Randomized splicing: A note on a simple technique for masking speech content. *Journal of Experimental Research in Personality,* 1971, *5,* 155–159.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 1977, *32,* 752–760.

Staffieri, J. R., & Bassett, J. E. Birth order and perception of facial expressions. *Perceptual and Motor Skills,* 1970, *30,* 606.

Sweeney, M. A., & Cottle, W. C. Nonverbal acuity: A comparison of counselors and noncounselors. *Journal of Counseling Psychology,* 1976, *23,* 394–397.

Tagiuri, R. Person perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., Vol. 3). Reading, Mass.: Addison-Wesley, 1969.

Vinacke, W. E. The judgment of facial expressions by three national-racial groups in Hawaii: I. Caucasian faces. *Journal of Personality,* 1949, *17,* 407–429.

Vinacke, W. E., & Fong, R. W. The judgment of facial expressions by three national-racial groups in Hawaii: II. Oriental faces. *Journal of Social Psychology,* 1955, *41,* 185–195.

Walton, W. E. Empathic responses in children. *Psychological Monographs,* 1936, *48*(1, Whole No. 213).

Weisgerber, C. A. Accuracy in judging emotional expressions as related to college entrance test scores. *Journal of Social Psychology,* 1956, *44,* 233–239.

Weitz, S. (Ed.). *Nonverbal communication: Readings with commentary.* New York: Oxford University Press, 1974.

Zaidel, S. F., & Mehrabian, A. The ability to communicate and infer positive and negative attitudes facially and vocally. *Journal of Experimental Research in Personality,* 1969, *3,* 233–241.

Zuckerman, M., Hall, J. A., DeFrank, R. S., & Rosenthal, R. Encoding and decoding of spontaneous and posed facial expressions. *Journal of Personality and Social Psychology,* 1976, *34,* 966–977.

Zuckerman, M., Lipets, M. S., Koivumaki, J. H., & Rosenthal, R. Encoding and decoding nonverbal cues of emotion. *Journal of Personality and Social Psychology,* 1975, *32,* 1068–1076.

Received May 12, 1977 ■