

AD-A102 726 TEXAS CHRISTIAN UNIV FORT WORTH INST OF BEHAVIORAL R--ETC F/8 5/10  
AGGREGATION BIAS IN ESTIMATES OF PERCEPTUAL AGREEMENT.(U)  
UNCLASSIFIED AUG 81 L R JAMES N00014-80-C-0315  
IBR-81-12 NL

| OF |  
ADA  
100726

END  
DATE  
FILMED  
9-81  
DTIC



9. School of Psychology  
Georgia Institute of Technology  
Atlanta, Georgia 30332

20. individuals in the same environment tend to agree with respect to climate perceptions. The present paper demonstrates that Drexler's analysis provided inflated estimates of agreement among individuals. The logic of the approach is then extended to other studies in which inflated estimates of agreement appeared likely.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A	

Aggregation Bias in Estimates of  
Perceptual Agreement

Lawrence R. James  
Georgia Institute of Technology

## Aggregation Bias in Estimates of Perceptual Agreement

The extent to which individuals agree with respect to perceptions of various aspects of their work environments has been addressed in a number of climate and climate-related studies (cf. Bass, Valenzi, Farrow, & Solomon, 1975; Drexler, 1977; Gavin & Howe, 1975; Howe, 1977; James, Demaree, & Hater, 1980; Jones & James, 1979; Litwin & Stringer, 1968; Payne & Mansfield, 1973; Payne & Pheysey, 1971; Pritchard & Karasick, 1973; Schneider, 1972; Schneider & Bartlett, 1970; Schneider & Snyder, 1975; Campbell & Beaty, Note 1; Curtis, Note 2; Hater, Note 3). Reviews of these studies indicate that the range of estimates of perceptual agreement among individuals is .00 to .50, with a median of approximately .12 (James, Hater, Gent, & Bruni, 1978; James & Sells, in press; Jones & James, 1979; Hater, Note 3). These reviews were based on estimates of interrater reliability for a single rater (intraclass correlations) and proportions of variance in individuals' perceptions associated with variation among environments (eta-squares and omega-squares). Not included were estimates of reliabilities of mean perceptions per environment (e.g., Spearman-Brown corrected intraclass correlations--see Jones and James, 1979 for a discussion of this issue) and estimates subject to aggregation bias (see below).

An estimate of agreement at the higher end of the range of agreement values is an eta-square ( $\eta^2$ ) of .42 reported by

Drexler (1977). However, given the number of studies reporting much lower values, one would suspect that current reviews of climate would question the likelihood of perceptual agreement. Such is often not the case. Woodman and King (1978) considered as unresolved the question of what attributes (organizational versus individual) are measured by climate perceptions. The key study referenced to support the organizational attribute position was Drexler (1977). Landy and Trumbo (1980) went a step further and suggested that climate perceptions reflected sufficient agreement and consistency at the individual level to justify their use as descriptors of organizational climate. Drexler (1977) was the key supporting reference for perceptual consistency. Another case of selective attention, again based on Drexler (1977), was provided by Schneider, Parkington, and Buxton (1980, p. 254), who stated, "The assumption of agreement in perceptions has been demonstrated empirically and allows for the aggregation of data within settings, facilitating studies across settings (Drexler, 1977)."

Perhaps the fascination with the Drexler article stems from the fact that it was based on a large sample of individuals and groups (6,996 individuals, 1,256 workgroups) from 21 diverse organizations. It is unfortunate, therefore, that the reported  $\eta^2$  of .42 was subject to an aggregation bias, which suggests that conclusions drawn by Drexler and others regarding perceptual agreement and consistency are based on an inflated estimate of variance in individuals' perceptions accounted for by organiza-

tions. The initial objective of this article is to demonstrate how an aggregation bias led to an inflated  $\eta^2$  in the Drexler analysis. Moreover, inasmuch as aggregates are all too often used to estimate perceptual agreement, the inflation of agreement estimates resulting from aggregation bias is illustrated in other climate and nonclimate studies.

#### Aggregation Bias in the Drexler Study

Drexler (1977) concluded not only that "42.2% of the variance in climate could be accounted for by organization" (p. 40), but also that James and Jones' (1974) use of the term "psychological climate" is "misleading if it connotes a construct that is largely intraindividual" (p. 41, italics added). These conclusions would lead one to believe that 42% of the variance in individuals' climate perceptions had been accounted for by differences in the 21 organizations. However, Drexler did not employ individuals' climate perceptions as the dependent variable. The dependent variable was mean perceptual scores per workgroup, which is to say that the  $\eta^2$  of .42 was based on an experimental design that employed  $K = 21$  organizations as the independent variable and 1,256 group means, nested within 21 organizations, as scores on the dependent variable. Consequently, interpretation of the  $\eta^2$  of .42 as if it had been calculated on individuals' perceptions almost assuredly provided an inflated estimate of agreement at the individual level. As discussed below, this is a form of aggregation bias known as the "ecological fallacy"

(cf. Hannan, 1971, 1973; Roberts, Hulin, & Rousseau, 1978; Robinson, 1950).

An ecological fallacy occurs when relationships, or functions of relationship indicators such as  $\eta^2$ , among individual level data are inferred from relationships among calculated aggregates of individual level data. Typically, relationships among aggregates provide inflated, technically spurious (cf. Hannan, 1971), estimates of relationships among individual level data. This is easily shown statistically for the Drexler data. Consider first the following three variance terms: (a)  $\sigma_I^2$  -- the variation of the 6,996 individual perceptions about the grand-mean of all individuals' perceptions ( $\underline{G}$ ), (b)  $\sigma_{wm}^2$  -- the variation of the 1,256 mean workgroup scores about  $\underline{G}$ , and (c)  $\sigma_o^2$  -- the variation of the mean organizational scores about  $\underline{G}$ .<sup>1</sup> We may now derive three  $\eta^2$ s, namely: (a)  $\eta_1^2 = \sigma_o^2 / \sigma_{wm}^2$  -- the proportion of variance in mean workgroup scores accounted for by differences in organizations; (b)  $\eta_2^2 = \sigma_o^2 / \sigma_I^2$  -- the proportion of variance in individuals' perceptions accounted for by differences in organizations; and (c)  $\eta_3^2 = \sigma_{wm}^2 / \sigma_I^2$  -- the proportion of variance in individuals' perceptions accounted for by differences in workgroups.

An estimator of agreement among individuals is  $\eta_2^2$ . Drexler's estimate of .42 was, however, predicated on  $\eta_1^2$ . The potential for an ecological fallacy is made evident by algebraic derivation, which shows that  $\eta_2^2 = \eta_1^2 \eta_3^2$ . This equation suggests that the appropriate estimate of perceptual agreement ( $\eta_2^2$ ) will



be equal to the Drexler estimate ( $\underline{\eta}_1^2$ ) only in the condition that 100% of the variance in individuals' perceptions is accounted for by differences in workgroups (i.e.,  $\underline{\eta}_3^2 = 1.0$ ). If  $\underline{\eta}_3^2$  is less than 1.0, then it follows that  $\underline{\eta}_2^2$  is less than  $\underline{\eta}_1^2$ , and the Drexler approach provides an inflated estimate of agreement at the individual level.

The degree of bias in the Drexler estimate can only be ascertained by a reanalysis of the data (i.e., compute  $\underline{\eta}_2^2$  instead of  $\underline{\eta}_1^2$ ). It is assumed that the bias would be sizeable. This assumption is based on a study by Bass et al. (1975),

which reported the highest levels of  $\underline{\eta}_3^2$  observed recently. The largest reported value of  $\underline{\eta}_3^2$  for an organizational level variable was .504 (external environment in the library directory sample--see Table 3). The median value of  $\underline{\eta}_3^2$  for organizational level variables was .325. One might now extrapolate, and as a heuristic exercise insert these values as estimates in the equation  $\underline{\eta}_2^2 = \underline{\eta}_1^2 \underline{\eta}_3^2$ . With  $\underline{\eta}_1^2 = .422$  (the Drexler value), we find  $\underline{\eta}_2^2$  equal to .213 and .137 for the highest value (.504) and the median value (.325) of  $\underline{\eta}_3^2$ , respectively.

Interestingly, these estimates, particularly the latter, are in line with the median value of perceptual agreement found in the reviews by Hater (Note 3), James et al. (1978), James and Sells (in press), and Jones and James (1979).

The statistics above were based on  $\underline{\eta}^2$  because Drexler employed  $\underline{\eta}^2$ . A reviewer suggested another approach for

demonstrating aggregation bias. This approach consists of viewing  $\eta_1^2$  (.42) as a reliability of mean scores per workgroup resulting from an upward adjustment in an intraclass correlation for the average number of raters per workgroup. The logic here is that the intraclass correlation (ICC) is an estimate of interrater reliability at the level of the individual rater (cf. Shrout & Fleiss, 1979), and can be estimated by the Spearman-Brown (SB) prophecy equation given knowledge of the reliability of group means and the average number of individuals per workgroup ( $5.57 = 6,996 \text{ individuals}/1,256 \text{ workgroups}$ ). The equation is  $.42 = 5.57 \text{ ICC}/(1 + 4.47 \text{ ICC})$ ; the resulting estimate of ICC is .12. This estimate is about the same as the .137 suggested above for  $\eta_2^2$  based on the median value of  $\eta_3^2$ .

In conclusion, it appears that the Drexler (1977) results were subject to an aggregation bias that, when corrected, is consistent with other studies of perceptual agreement. Hopefully, this will stimulate some current reviewers of organizational climate to reconsider their conclusions regarding perceptual agreement, perhaps by broadening the scope of their reviews to include not only other climate studies but also the reasons that individuals in the same organization might cognitively construct somewhat different perceptions (cf. Ekehammar, 1974; James & Jones, 1976; James et al., 1978; Payne & Mansfield, 1973).

#### Aggregation Bias in Other Studies of Perceptual Agreement

Estimates of agreement based on group mean scores have been

incorrectly interpreted as applying to individuals' perceptions in a number of other studies. Specific reference is directed to studies of perceptual agreement among members of different roles (e.g., supervisors-subordinates, incumbents-observers, customers-employees) in which either (a) the perception of a member of one role (e.g., a supervisor) is correlated with the aggregate perception of two or more members of another role (e.g., subordinates) on a sample of K role sets, or (b) two sets of aggregate perceptions are correlated (e.g., mean customer scores and mean employee scores on a sample of K organizations). The former procedure is referred to as the "single aggregate approach", and the latter the "double aggregate approach." Each approach is discussed briefly below.

Single aggregate approach. This approach is illustrated by the following studies: (a) Evans (1972), who measured agreement among perceptions of leader behavior by correlating supervisors' self-descriptions with means of subordinates' descriptions (using groups as observations); (b) Oldham (1976), where agreement was assessed by correlating focal managers' perceptions of their own motivational strategies with means of subordinates' perceptions; and (c) Schneider (1972) and Schneider and Bartlett (1970), who estimated "agreement on climate perceptions across positions" by correlating the mean climate perceptions of agents with the climate perceptions of agency managers on a sample of life insurance agencies. While none of these studies reported particularly high levels of agreement, it is nevertheless likely

that the estimates were inflated. This is because the use of aggregates (means) of subordinates' (agents') perceptions deleted from the analysis within-group (within-agency) variance in subordinates' (agents') perceptions. That is, of course, error variance.

The likelihood of inflated estimates of agreement is demonstrated statistically by employing an analogue of an equation presented by James et al. (1980, Eq. 5) to assess relationships between situational variables and individual variables. For example, if the  $i^{\text{th}}$  manager's climate score ( $Y_{ik}$ ) is assigned to all agents in the  $k^{\text{th}}$  agency ( $k=1, \dots, K$  agencies), then the correlation between managers' climate scores and agents' climate scores ( $X_{jk}$  --  $j=1, \dots, n_j$  agents per agency), based on all agents across the sample of  $K$  agencies, takes the form:  $r_{yx} = \eta_k r_{y\bar{x}}$ .<sup>2</sup> The correlation  $r_{yx}$  is an estimate of agreement between managers and agents at the individual level of analysis (i.e., no scores have been aggregated), the square of  $\eta_k$  reflects the proportion of variance in agents' climate scores accounted for by differences in the  $K$  agencies, and  $r_{y\bar{x}}$  is the correlation between managers' climate scores and mean climate scores for agents, based on the total agent sample. Given equal  $n_j$ , the latter correlation,  $r_{y\bar{x}}$ , provides the same value as the statistic used by Evans (1972), Oldham (1976), Schneider (1970), and Schneider and Bartlett (1970) to compute estimates of perceptual agreement. Note that  $r_{y\bar{x}}$  will be equal to  $r_{yx}$  only in the condition that  $\eta_k = 1.0$ , which suggests that all agents in each

agency agreed perfectly (i.e., there is no within-agency variance in agents' climate scores).

It is extremely unlikely that an  $\eta_k$  will equal 1.0. Consequently, given that  $\eta_k$  is less than 1.0, it follows that  $r_{yx}$  is less than  $r_{y\bar{x}}$ . Thus,  $r_{y\bar{x}}$  provides an inflated estimate of perceptual agreement at the individual level of analysis.

Double aggregate approach. Examples of this approach are seen in the following studies: (a) Hackman and Lawler (1971), Hackman and Oldham (1975), Hackman, Pearce, and Wolfe (1978), and Oldham, Hackman, and Pearce (1976), who, for example, assessed agreement between supervisors' and subordinates' perceptions of a job dimension by correlating mean supervisory perceptions with mean subordinate perceptions, using jobs as the sample;<sup>3</sup> (b) Ilgen & Fugii (1976), who, after showing that observers and group members did not agree on descriptions of leader behavior at the individual level of analysis, proceeded to reestimate agreement by computing correlations between mean observer perceptions and mean group member perceptions on a sample of groups; and (c) Schneider and Snyder (1975), who correlated means of managers' climate perceptions and means of trainees' climate perceptions on a sample of life insurance agencies to test the hypothesis that "people in an organization should agree more on their description of the climate than on their feelings of job satisfaction" (p. 319, italics added to emphasize that the level of interpretation is individuals).

Basing estimates of agreement on double aggregates is an

exacerbation of the problem with single aggregates. That is, rather than deleting within-group (within-job, within-agency) error variance for one set of groups, it is now being deleted for two sets of groups. Thus, unless the  $\bar{n}^2$  for each group is equal to 1.0 (i.e., there is no within-group variation for either group), a correlation of aggregates will provide an inflated estimate of agreement among individuals.

The substance, although not the precise form, of the statistical bias resulting from using double aggregates to estimate perceptual agreement among individuals is illustrated in the following scenario. The basic question addressed here is whether individuals who rated the same job agreed. Suppose that we have a sample of 100 jobs and 10 different raters for each job, where the raters may be job incumbents, supervisors, observers, etc. An estimate of perceptual agreement among individuals is then calculated using the ICC procedure to provide an interrater reliability at the level of the individual rater.<sup>4</sup> Now suppose that we (a) randomly split the 10 raters for each job into two groups of five, (b) calculate a mean for each group of five, and (c) correlate the means using the 100 jobs as the sample. If the ICC is arbitrarily set at .30, then the correlation among means may be estimated by applying the SB equation, using a correction factor of five (i.e., five scores per mean). The resulting value is .68, but then this is an estimate of the reliability of means and a highly inflated estimate of agreement at the individual level.

Conclusions

The moral of the story is simple. If one is assessing perceptual agreement among individuals, then the appropriate level of analysis is the individual. Furthermore, to avoid misunderstanding, the discussion here was limited to the fallacy of interpreting agreement estimates based on aggregates as applying to agreement among individuals. The use of aggregates for other purposes was not addressed. What justifies the calculation of an aggregate, making sense out of what is measured by an aggregate, and interpreting relationships among aggregates are subjects that require considerable discussion and are prone to debate. Compare, for example, Ilgin and Fugii's (1976) and Katona's (1979) justifications for aggregate level analysis with issues raised by Firebaugh (1978, 1980) concerning what is measured by an aggregate and interpretation of relationships among aggregates.

References

- Bass, B. M., Valenzi, E. R., Farrow, D. L., & Solomon, R. J. Management styles associated with organizational, task, personal, and interpersonal contingencies. Journal of Applied Psychology, 1975, 60, 720-729.
- Drexler, J. A., Jr. Organizational climate: Its homogeneity within organizations. Journal of Applied Psychology, 1977, 62, 38-42.
- Ekehammer, B. Interactionism in personality from a historical perspective. Psychological Bulletin, 1974, 81, 1026-1048.
- Evans, M. G. Leadership behavior: Demographic factors and agreement between subordinate and self-descriptions. Personnel Psychology, 1972, 25, 649-653.
- Firebaugh, G. A rule for inferring individual-level relationships from aggregate data. American Sociological Review, 1978, 43, 557-572.
- Firebaugh, G. Assessing group effects: A comparison of two methods. In E. F. Borgatta and D. J. Jackson (Eds.), Aggregate data: Analysis and interpretation. Beverly Hills, CA: Sage Publications, 1980.
- Gavin, J. R., & Howe, J. G. Psychological climate: Some theoretical and empirical considerations. Behavioral Science, 1975, 20, 228-240.



- Hackman, J. R., & Lawler, E. E., III. Employee reactions to job characteristics. Journal of Applied Psychology Monograph, 1971, 55, 259-286.
- Hackman, J. R., & Oldham, G. R. Development of the job diagnostic survey. Journal of Applied Psychology, 1975, 60, 159-170.
- Hackman, J.R., Pearce, J. L., & Wolfe, J. C. Effects of changes in job characteristics on work attitudes and behaviors: A naturally occurring quasi-experiment. Organizational Behavior and Human Performance, 1978, 21, 289-304.
- Hannan, M. T. Problems of aggregation. In H. M. Blalock (Ed.), Causal models in the social sciences. Chicago: Aldine, 1971.
- Hannan, M. T. Aggregation and disaggregation in sociology. Lexington, MS: Lexington Books, 1973.
- Howe, J. G. Group climate: An exploratory analysis of construct validity. Organizational Behavior and Human Performance, 1977, 19, 106-125.
- Ilgen, D. R., & Fugii, D. S. An investigation of the validity of leader behavior descriptions obtained from subordinates. Journal of Applied Psychology, 1976, 61, 642-651.
- James, L. R., Demaree, R. G., & Hater, J. J. A statistical rationale for relating situational variables and individual differences. Organizational Behavior and Human Performance, 1980, 25, 354-364.
- James, L. R., Hater, J. J., Gent, M. J., & Bruni, J. R. Psychological climate: Implications from cognitive social learning theory and interactional psychology. Personnel

Psychology, 1978, 31, 781-813.

James, L. R., & Jones, A. P. Organizational climate: A review of theory and research. Psychological Bulletin, 1974, 81, 1096-1112.

James, L. R., & Jones, A. P. Organizational structure: A review of structural dimensions and their conceptual relationships with attitudes and behavior. Organizational Behavior and Human Performance, 1976, 16, 74-113.

James, L. R., & Sells, S. B. Psychological climate: Theoretical perspectives and empirical research. In D. Magnusson (Ed.), Toward a psychology of situations: An interactional perspective. Hillsdale, NJ: Lawrence Erlbaum, in press.

Jones, A. P., & James, L. R. Psychological climate: Dimensions and relationships of individual and aggregated work environment perceptions. Organizational Behavior and Human Performance, 1979, 23, 201-250.

Katona, G. Toward a macropsychology. American Psychologist, 1979, 34, 118-126.

Landy, F. J., & Trumbo, D. A. Psychology of work behavior. Homewood, IL: Dorsey Press, 1980.

Litwin, G., & Stringer, R. Motivation and organizational climate. Boston: Harvard University Press, 1968.

Oldham, G. R. The motivational strategies used by supervisors: Relationships to effectiveness indicators. Organizational Behavior and Human Performance, 1976, 15, 66-86.

- Oldham, G. R., Hackman, J. R., & Pearce, J. L. Conditions under which employees respond positively to enriched work. Journal of Applied Psychology, 1976, 61, 395-403.
- Payne, R. L., & Mansfield, R. Relationships of perceptions of organizational climate to organizational structure, context, and hierarchical position. Administrative Science Quarterly, 1973, 18, 515-526.
- Payne, R. L., & Pheysey, D. C. G. G. Stern's organizational climate index: A reconceptualization and application to business organizations. Organizational Behavior and Human Performance, 1971, 6, 77-98.
- Pritchard, R. D., & Karasick, B. W. The effect of organizational climate on managerial job performance and job satisfaction. Organizational Behavior and Human Performance, 1973, 9, 126-146.
- Robinson, W. S. Ecological correlations and the behavior of individuals. American Sociological Review, 1950, 15, 351-357.
- Roberts, K. H., Hulin, C. L., & Rousseau, D. M. Developing an interdisciplinary science of organizations. San Francisco: Jossey-Bass, 1978.
- Schneider, B. Organizational climate: Individual preference and organizational realities. Journal of Applied Psychology, 1972, 56, 211-217.
- Schneider, B., & Bartlett, C. J. Individual differences and organizational climate II: Measurement of organizational

climate by the multi-trait, multi-rater matrix. Personnel Psychology, 1970, 23, 493-512.

Schneider, B., Parkington, J. J., & Buxton, V. M. Employee and customer perceptions of service in banks. Administrative Science Quarterly, 1980, 25, 252-267.

Schneider, B. & Snyder, R. Some relationships between job satisfaction and organizational climate. Journal of Applied Psychology, 1975, 60, 318-328.

Shrout, P. E., & Fleiss, J. L. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-428.

Woodman, R. W., & King, D. C. Organizational climate: Science or folklore? Academy of Management Review, 1978, 3, 816-826.

Reference Notes

1. Campbell, J. P. & Beaty, E. E. Organizational climate: Its measurement and relationship to work group performance. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September, 1971.
2. Curtis, B. Boundary conditions for relationships among components of job perceptions and attitudes. Unpublished dissertation, Texas Christian University, 1976.
3. Hater, J. J. Agreement among perceptions of psychological climate: A comparison of within-group and between-group designs. Unpublished thesis, Texas Christian University, 1977.

Footnotes

Support for this project was provided under Office of Naval Research Contract N00014-80-C-0315, Office of Naval Research Project NR170-904. Opinions expressed are those of the author and are not to be construed as necessarily reflecting the official view or endorsement of the Department of the Navy.

The author wishes to thank Robert G. Demaree, Robert L. Ellison, Charles V. Riche, S. B. Sells, Gerrit Wolf, and C. Michael York for their helpful suggestions and advice.

Requests for reprints should be sent to Lawrence R. James, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.

<sup>1</sup>To simplify statistical developments, population parameters were employed and it was assumed that (a) the number of individuals in each workgroup was the same for all workgroups, and (b) the number of workgroups in each organization was equal for all organizations. Thus, for example, the grand mean of individuals' scores was the same as the grand mean of mean workgroup scores, and the means per organization of individuals' scores and mean workgroup scores were equivalent. While unrealistic in practice, these assumptions do not affect the logic and conclusions of the statistical critique.

<sup>2</sup>See James et al. (1980) for assumptions (e.g., linearity) to interpret this statistic. In the present application, it was

not assumed that the  $Y_{ik}$  was homogeneous with respect to agents, although the  $n_j$  were considered equal.

<sup>3</sup>Hackman and Oldham (1975, p. 164) referred to correlations among means involving employees and both supervisors and researchers as "indirect tests of the 'objectivity' of employee ratings" rather than tests of agreement. It is also the case that (a) these correlations were compared to estimates of agreement in the Hackman and Lawler (1971) study, (b) Hackman and Lawler (1971, p. 268) stated that it is not "possible to demonstrate conclusively that employee judgments are objectively accurate, because no unambiguous standard of accuracy is available", and (c) later studies (Hackman et al., 1978; Oldham et al., (1976) returned to the use of the term "agreement." It appeared reasonable, therefore, to regard the Hackman and Oldham (1975) estimates as agreement indices.

<sup>4</sup>The statistical model employed in this scenario involves a random effects, one-way ANOVA and the ICC equation for incomplete designs (cf. Shrout & Fleiss, 1979). A separate component for type of rater (e.g., supervisor, subordinate) could be included by using a more sophisticated design, such as a general linear model with dummy variables to represent jobs and types of raters, accompanied by appropriate interaction terms. Nevertheless, the basic question is whether individuals rating the same job agreed, which is the question addressed by the simple design.

ATE  
LMED  
-8