failed to specify the complete history necessary for equivalence patterns to emerge. I think it could be safely argued that the naming hypothesis begs the same question.

To conclude, Horne and Lowe have offered an interesting, novel, and in-depth look at how naming relations might be established from more basic verbal units. The developmental analysis is sure to spark research ideas, and thus may fulfill certain goals of the paper. With respect to accounts of stimulus equivalence, however, I remain unconvinced that the naming hypothesis provides a significant departure from or improvement on the alternatives. Incomplete facets of a theory can always be developed further, but I particularly question the ability of the suggested research agenda to further a data-based resolution of equivalence accounts, given the difficulties of potential disconfirmation outlined here.

METHODOLOGICAL ISSUES IN THE STUDY OF NAMING

MARK GALIZIO

UNIVERSITY OF NORTH CAROLINA AT WILMINGTON

Horne and Lowe provide a well-developed account of the acquisition of the naming relation and an interpretation of its role in determining various forms of verbal and symbolic control of behavior. Their account is rich and includes many controversial points, so there is much more that warrants attention than can be discussed in any single commentary. Accordingly, I will leave debate of many of the conceptual and theoretical issues to others, and focus on the parts of Horne and Lowe's article that I believe may provide direction for much of the research that is sure to be generated.

Horne and Lowe outline three research areas that they take to be key tests of the naming hypothesis: (a) equivalence in nonhumans, (b) equivalence in nonverbal humans, and (c) naming manipulations and equivalence in verbal humans. Although they recognize that "much of the critical experimentation remains to be done" (p. 240), they clearly contend that the available evidence supports the naming hypothesis. I will review these three areas here, with the aim of considering methodological problems that have made critical experiments difficult to conduct. In my view, the currently available studies are insufficient to differentially support Horne and Lowe's position relative to the rival hypotheses. Research strategies that might clarify these issues will be considered.

Equivalence in Nonhumans

Certainly any theory must account for the observation that nonhuman organisms generally fail the tests of stimulus equivalence on which verbal human subjects so easily succeed. For Horne and Lowe, this state of affairs is a predictable consequence of nonhumans' lack of naming skills. On the other hand, accounts of the nonhuman data can also be developed from Sidman (1994) and Hayes (1991), so that the simple failure of nonhuman subjects' performances to show symmetry and transitivity after conditional discrimination training may not differentiate these theories. Nevertheless, because the

I would like to thank Katherine Bruce for helpful comments on the manuscript and Carol Pilgrim and the equivalence lab group for many stimulating discussions of these issues.

Address correspondence to Mark Galizio, Department of Psychology, University of North Carolina at Wilmington, 601 S. College Rd., Wilmington, North Carolina 28403 (E-mail: Galizio@uncwil.edu).

three accounts are quite different, they might be differentiated on the basis of how well they handle nonhuman success on equivalence tests. For example, in Schusterman and Kastak's (1993) study of stimulus equivalence in the sea lion, responses under the control of symmetry and transitivity relations were reinforced on some (but not all) trial types. The demonstration of control by equivalence relations on trial types for which reinforcement was unavailable may be attributed to the successful training of generalized equivalencing, thus supporting Hayes' relational frame theory. However, as Horne and Lowe point out, enthusiasm for such an interpretation is dimmed by Dugdale and Lowe's (1990) studies with chimpanzees, in which comparable manipulations were not successful in bringing about successful equivalence probe performances.

Another successful strategy has been to train distinctive response patterns to the training stimuli (McIntire, Cleary, & Thompson, 1987; Manabe, Kawashima, & Staddon, 1995). Even though the distinctive response in the experiment by Manabe et al. was vocal, the "naming" trained in these studies lacks the rich history of typical naming relations in humans, as characterized by Horne and Lowe, and thus the success of such training is more consistent with other accounts (see Sidman, 1994, pp. 466–470). Any conclusions in this key area are clearly premature, but because the three theoretical approaches have different accounts of the prerequisites of equivalence test performance, each suggests different strategies for training nonhumans to pass equivalence tests. Thus, there are reasonable prospects for critical tests from the animal laboratory (with due caution regarding the formalistic fallacy).

Equivalence in Nonverbal Humans

From Horne and Lowe's perspective, humans who lack naming skills would also be expected to fail equivalence tests. Despite the importance of this issue, there is virtually no compelling evidence that links equivalence failure to the absence of naming skills. One problem with studies in the relevant literature is that nonverbal subjects may be handicapped in other respects that affect match-to-sample performance in the typical human testing situation. Thus, a major methodological difficulty is the problem of matching verbal and nonverbal subjects with respect to general learning ability and motivational variables. In addition, social contact with the experimenter and other features of the experimental procedure may differentially affect nonverbal subjects. An example is the practice of having the experimenter provide assurance when subjects ask questions about the experiment (Barnes, McCullagh, & Keenan, 1990; Devany, Hayes, & Nelson, 1986). Such procedures inevitably result in differential treatment of verbal and nonverbal subjects, and may affect performances in ways that are not easily interpreted. Thus, these studies have not convincingly demonstrated that failure to pass equivalence tests in nonverbal subjects is related to the subjects' inability to name the stimuli.

Even disregarding the many possible sources of confounding effects in comparisons of verbal and nonverbal subjects, the published literature does not provide compelling evidence that naming is critical for the emergence of equivalence. Consider the Devany et al. (1986) study that is often cited as showing that normally developing and mentally handicapped children with language show equivalence, whereas mentally handicapped children without language do not. In that study, acquisition of the prerequisite conditional discriminations was much slower in each of the children without language, and the failure of equivalence to emerge in a single 40-trial test in these subjects may have been merely another reflection of their delayed acquisition. Systematic replications of this study with more extended and conventional testing conditions are badly needed. Perhaps one reason for the limited research in this area is the surprising difficulty researchers have had in training arbitrary match to sample in young children without special prompting (e.g., Augustson & Dougher, 1992). In any case, there is potential for critical research in this area, both with preverbal normally developing children and with subjects who lack naming skills for other reasons, although interpretation of negative results is always problematic. Thus, the kind of research that will be required in order for such studies to be compelling are systematic, intensive, and extended explorations, with nonverbal subjects, of training and testing

variables that have been shown to modulate equivalence phenomena. Support for Horne and Lowe's view would be provided if systematic replications were consistent in showing that successful performances on equivalence tests fail to emerge in nonverbal or preverbal subjects in a selective fashion.

Naming Manipulations and Equivalence

Horne and Lowe discuss studies in which normally developing (Dugdale & Lowe, 1990) or autistic (Eikeseth & Smith, 1992) children who initially fail equivalence tests show improved performance when they are trained to produce a common name to the stimuli within classes. Although these findings are certainly of interest, they remain open to various interpretations. One problem with these studies is that training of naming is confounded with extent of overall exposure to the problem. In addition, training of naming may produce effects other than those specifically involving the naming relation. For example, such training may induce differential attention to the test stimuli (see Galizio & Baron, 1976), it may serve an "instructional" role, or it may generate functional equivalence; any of these effects could alter equivalence test performance in one direction or another (e.g., Dickins, Bentall, & Smith, 1993). Given the inclusion of appropriate control conditions, however, studies in this area promise to extend the experimental analysis of verbal behavior in important ways. But once again, because the rival hypotheses also acknowledge that naming and other verbal manipulations (e.g., instructions) may influence match-to-sample performance, experiments that provide critical tests may not be possible.

Beyond Match to Sample

This brief evaluation of relevant research might be summarized as finding no convincing evidence either against Horne and Lowe's approach or favoring it. The only clear conclusion at this stage is that the defining experiments in these areas have yet to be conducted. In its favor, however, perhaps Horne and Lowe's analysis may lead behavior analysts to another potential key test area: research that directly extends these analyses to the study of natural language classes and the phenomena associated with them. One important reason for current interest in equivalence research is that it has seemed to provide a useful model for analysis of natural language categories and related verbal behavior. Arbitrary match-to-sample procedures have provided the framework for most of the thinking in this area and remain the defining paradigm. Nevertheless, the application of these procedures to the many kinds of naturally occurring behavior that may be interpreted in terms of equivalence-like processes is sometimes awkward. Horne and Lowe's conception of the naming relation is neither developed nor defined in terms of the match-to-sample procedure, and although this feature has the disadvantage of losing the rigor of that well-elaborated method, it may lend itself more readily to analysis of natural language phenomena.

One final comment seems to be appropriate regarding the controversies that are bound to be generated by the commentary-and-reply format used here. Although substantial disagreements are present and persistent, they should not overshadow the enormous progress that has already been made in this field. Research on stimulus equivalence and related areas has begun to generate successful behavior-analytic explanations for a host of problems that have been viewed by critics from traditional psychology as beyond the scope of behavior analysis.