

Expert and novice intuitive judgments about animal behavior

MICHAEL J. RENNER

West Chester University, West Chester, Pennsylvania

and

CATHERINE HACKETT RENNER

Zeneca Pharmaceuticals, Wilmington, Delaware

Although research in comparative psychology often relies on direct behavioral observation for data gathering, the influence of the observer is often overlooked. To study the effect of expertise on making judgments about behavior, novices and experts observed videotapes of rats from enriched and impoverished environments and attempted to determine each rat's history. The judgment accuracy of all groups exceeded chance. Although experts used a broader vocabulary for describing the foundations of their judgments concerning animal behavior, novices were as accurate as experts at making the discrimination. Subjects were not able to correctly assess their own accuracy. The results of this study suggest that observer expertise does influence the data-gathering process in comparative psychology, but in indirect and often subtle ways.

Because much of the data in the field of comparative psychology is gathered by human observers who make judgments in real time concerning the classification of their subjects' behavior (cf. Altmann, 1974), the issue of rater reliability is often addressed in treatments of observational research methods (see, e.g., Martin & Bateson, 1986) and has been directly studied. However, the issue of the expertise of the judge remains essentially unaddressed. Because these real-time ratings of behavior are made quickly and necessarily involve some element of subjective judgment, it is reasonable to speculate that a certain amount of inaccuracy occurs.

Errors in behavior ratings may or may not be diminished by the level of expertise of the judge. It is naturally assumed that there is a difference between experts and novices in problem-solving ability. Early studies of expertise in several domains (chess masters, Chase & Simon, 1973; swine judges, Phelps & Shanteau, 1978; mathematics, de Sessa, 1982; and physics, Larkin, McDermott, Simon, & Simon, 1980) converged on two generalizations: first, that experts and novices use similar cognitive processes in solving problems, and second, that they differ in the nature of the knowledge base on which they draw for decision making. Experts' knowledge is well organized and richly interconnected, which allows for rapid problem assessment, identification of critical features, and causal analysis of problems. The problem-solving strategies of novices tend to be weak all-purpose methods that do not require much knowledge about the domain.

In the present study, human observers were asked to perform a discrimination involving judgments in real time about individual animals' experience histories. There is a large body of research literature concerning the effects of housing environment on behavior, neural anatomy, and function (reviewed in Renner & Rosenzweig, 1987). Briefly, these studies indicate that rats housed in a relatively enriched environment (EC) display increases in the size and complexity of many neocortical features, as well as altered neurochemical activity in the neocortex, and that they are better able to solve complex behavioral problems than are littermates housed in relatively impoverished conditions (IC). It has been shown that EC rats organize their behavior differently from IC rats' while exploring a neutral environment unlike either housing condition (Renner, 1987). In addition, it is common for researchers who work on problems that involve manipulating rats' experience histories to say that behaviors of EC and IC animals differ recognizably.

The present experiment was designed to determine whether the expertise of observers would apply to an unfamiliar problem, and secondarily, whether human observers could discriminate EC rats from IC rats in a neutral situation without specified criteria. We used two groups of rats that differed in their recent housing conditions, videotaped in a laboratory situation. These videotapes provided the basis for an earlier report (Renner, 1987) showing differences between these groups in gross locomotor exploration and the investigation of inanimate nonfood stimulus objects.

METHOD

Subjects

Three groups of human observers ($N = 40$) participated as subjects in this study. Two groups were made up of novices ($n_s = 12$ and 11)—

Correspondence concerning this manuscript may be addressed to M. J. Renner, Department of Psychology, West Chester University, West Chester, PA 19383 (e-mail: mrenner@wcu.bitnet).

undergraduate psychology majors taking an upper division course in experimental design, who had no formal experience or training in observational techniques or animal behavior. The third group (*experts*, $n = 17$) comprised advanced students and faculty in a biological psychology research group. Most, but not all, members of this group specialized in animal behavior rather than behavioral neuroscience, and their work involved direct observation that shared some characteristics with the task presented here. None of the observers from any group had studied this problem previously. Because of the somewhat variable nature of their backgrounds, after all other procedures reported here had been completed, the experts were asked to provide a self-assessment (on a scale from 1 to 5) of their degrees of experience with direct behavioral observation.

Procedure

The subjects were first presented a brief videotape segment, showing rats in the enriched and impoverished environments. Next, they were shown twenty 1-min videotape segments of individual animals presented in random order in a neutral context. (This context, an arena unlike either housing condition, provided the animal with a free choice of activities; it is described in detail in Renner, 1987.) The subjects were instructed to indicate, on a response sheet, which of the two possible environments they believed a particular animal to be from, and how confident they were of this decision (with 50% identified as "pure guessing" and 100% as "absolutely sure"). Half of the 20 rats shown in the videotape segments came from each of the different environments, although the subjects were not informed of this. Each 1-min videotape segment was followed by a 15-sec blackout period, to allow the subjects time for recording their judgment and confidence ratings for that segment. For one group of novices, the behavior sample was chosen arbitrarily; the 4th min of a 10-min session in the arena was used. Because this group reported that activity level was the primary basis for their discriminations, the other groups (the expert group and the second group of novices) were shown a different set of tape segments, which were selected to control for activity level between EC and IC rats. These tape segments were taken from the same 10-min arena sessions, but were created by selecting the 1st min of the session during which the subject was active in the arena for more than half the minute. After completing these judgments, the subjects were asked to provide written narratives explaining the basis of their discriminations between groups.

RESULTS

The judgment accuracy of all groups significantly exceeded chance [$t(11) = 7.43$, $p < .001$, for the novice group watching the arbitrarily selected tape; $t(10) = 12.86$, $p < .001$, for the novice group watching the activity-controlled tape; $t(16) = 10.52$, $p < .001$, for the expert group, also watching the activity-controlled tape]. However, experts were no more successful at discriminating EC from IC animals (73% correct) than novices (72% correct for the arbitrarily selected tape group, and 76% correct for the activity-controlled tape group). There were significant differences in the narratives provided by the subjects. Experts' descriptions included watching for a higher number of criteria [$t(26) = 2.04$, $p < .05$]. Subjectively, it also seemed that experts provided more concrete detail in their narratives than did the novices. The correlation between self-ratings of expertise by the subjects in the expert group and their actual judgment accuracy was nonsignificant ($r = .18$).

In addition to finding no evidence that expertise in a broad sense aided subjects in making the judgments required here, no strong evidence was found to indicate that individuals in any group could accurately make a more circumscribed assessment of their own performance on

this particular task. (In the judgment literature, this is referred to as *calibration*. A well-calibrated observer would correctly estimate the accuracy of the behavioral judgments he/she was making.) For the expert group, the correlation between confidence in judgments and actual judgment accuracy was nonsignificant ($r = .13$). For novices, the calibration among those viewing the activity-controlled tape ($r = .37$) was also nonsignificant, and among those viewing the arbitrarily selected tape, only marginally significant ($r = .48$, $p = .05$). Of the 12 observers in this arbitrarily selected tape group, however, one was clearly an outlier (displaying confidence only slightly above chance, and actual judgment performance slightly below chance); dropping this subject eliminated any apparent calibration displayed by this group ($r = -.05$).

DISCUSSION

All three groups performed above chance at identifying the previous housing conditions of individual animals. It is therefore clear that there is some difference in the behaviors of rats with different histories of experience that is discriminable in real time by minimally instructed observers. Content analysis of the observers' self-reports do not, however, resolve the question of whether the difference between EC and IC detected by the judges is based on the same differences reported in Renner (1987) or on other, not-yet-quantified behavioral differences.

As long as the study of comparative psychology relies on observers' judgments for much of its data collection, factors that bear on the accuracy and validity of those judgments are legitimate and valuable topics for investigation within comparative psychology. The results of this study suggest that observers' expertise does influence the data-gathering process, but in indirect and often subtle ways. Although experts may have a broader vocabulary for describing the foundations of their judgments concerning animal behavior (or be otherwise better able to explain their judgments), the evidence presented here does not support the view that they are superior in the accuracy of their real-time judgments or in awareness of the accuracy of their own performance. If experts are better able to verbalize the criteria used for judgments, the influence that this increased ability might have on observational data gathering may not be straightforward, and it deserves closer scrutiny. Also, expert-novice differences may emerge in other tasks or more difficult tasks than the discrimination used in this study, but this possibility remains for future study.

REFERENCES

- ALTMANN, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, **49**, 227-267.
- CHASE, W. G., & SIMON, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.
- DE SESSA, A. A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science*, **6**, 37-75.
- LARKIN, J. H., McDERMOTT, J., SIMON, D. P., & SIMON, H. A. (1980). Models of competence and solving physics problems. *Cognitive Science*, **4**, 317-345.
- MARTIN, P., & BATESON, P. (1986). *Measuring behaviour*. Cambridge: Cambridge University Press.
- PHELPS, R. H., & SHANTEAU, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior & Human Performance*, **21**, 209-219.
- RENNER, M. J. (1987). Experience-dependent changes in exploratory behavior in the adult rat: Overall activity level and interactions with objects. *Journal of Comparative Psychology*, **101**, 94-100.
- RENNER, M. J., & ROSENZWEIG, M. R. (1987). *Enriched and impoverished environments: Effects on brain and behavior*. New York: Springer-Verlag.