

Observation Ability: Determining and Extending Its Presence

Stephen P. Norris

Institute for Educational Research and Development
Memorial University of Newfoundland

Ruth King

York University

Many arguments rely upon the truth of reports of observations. Arguments supporting one product over another often rely upon data containing results of scientific observation. The arguments of lawyers attempting to demonstrate the innocence of their clients, and of prosecutors trying to prove their guilt, often depend upon the testimony of eyewitnesses. Assessing such arguments requires the ability to judge correctly the believability of the observation reports on which they depend. Informal logic competence, conceived as competence in appraising arguments, thus involves the ability to judge reports of observations, an ability which is one aspect of observational ability (Norris, 1983). In this paper we are concerned with the degree to which people have this aspect of observational ability, and the manner of promoting it in them.

Observation reports, such as those offered in eyewitness testimony, are not perfectly reliable. This is a recognized problem for the courts (Loftus, 1979,

Ch. 10; Yarmey, 1979, Ch. 8) as well as for other fields. The situation is aggravated by yet another problem, which we document in fair detail in this paper: People are not adequately adept at distinguishing reliable observation reports from unreliable ones.

We assume that progress in this field relies fundamentally on specifying how the correct appraisal of observation reports is conducted. In addition, progress depends upon knowing the extent and nature of any lack of ability which people have in this area, and upon having a framework upon which instruction can be based. We briefly address each of these issues.

THE QUALITY OF PEOPLE'S APPRAISALS

In this first section we describe people's ability

to assess the believability of reports of observations. We use information from three sources: experiments conducted in the psychological study of eyewitness testimony; senior high school students' overall performances on a test of observational ability and their performances on the individual items of the test; and the thinking which a sample of senior high school students reported when asked what led them to choose the answers they did to items on the observation test. All the evidence points to the same conclusion: ability in this area is not as widespread as we would like. However, there do seem to be respects in which people are fairly competent, suggesting to us some reason for hope.

Trying Eyewitness Testimony

The most prevalent finding reported in the psychological literature on eyewitness testimony is that jurors place more confidence in eyewitness testimony than in any other type of courtroom evidence. In itself this degree of confidence would not be particularly troublesome. What is disturbing is that large proportions of people do not properly temper their confidence in such testimony in accord with factors which tend to reduce its accuracy. For example, results of experiments indicate that people are less accurate when identifying faces of those from other races than when identifying faces of their own race. However, from a sample of over five hundred adults, Loftus (1979) found that only about one-half of them were aware of this fact. The accuracy of eyewitness testimony is also adversely affected if the observer is under stress, though in the same study by Loftus only about one-third of the subjects realized this. Similarly, witnesses report upon violent events less reliably than upon non-violent events, but fewer than one-fifth of the subjects knew this. Finally, while victims confronted with a deadly weapon such as a gun or knife tend to focus their attention on that object, thereby reducing the accuracy of their reports about other features of the event, only about two-fifths of the five hundred adults recognized this accuracy-reducing effect.

Other research (Yarmey, 1979, for example) indicates further errors which jurors typically make in assessing eyewitness testimony. Jurors tend to judge testimony incorrectly according to such factors as witnesses' agreeableness on the witness stand their dress and their expressing confidence in the truth of what they testified. This latter factor, expressed degree of confidence, is in fact a very prominent one in the minds of jurors. Several studies (for example, Wells, Ferguson, and Lindsay, 1981; Lindsay, Wells, and Rumpel, 1981) suggest that the confidence of a witness is a primary determinant in jurors' judgments, even though confidence is only marginally related to accuracy according to the same

studies and others (Brown, Deffenbacher, and Sturgill, 1977; Leippe, Wells, and Ostrom, 1978; Wells, Leippe, and Ostrom, 1979). Fallacious thinking of this sort is of course particularly disquieting when a confident witness is one whose testimony has been affected by certain accuracy-reducing factors. Finally, people have been shown to base their judgment of testimony in accord with whether or not witnesses can remember minute details of events. For example, more credit is given to a witness' descriptions of the central features of an event, if that witness also reports on peripheral details such as colour of a culprit's shoes. Evidence suggests, however, that there is a **negative** relationship between memory of central and of peripheral aspects of an event (Haggen, Meacham, and Mesibov, 1970).

The conclusion usually drawn by researchers in the eyewitness testimony field is that in general people's ability to try testimony is weak. In addition, some specifics about the nature of that weakness emerge. Clearly, the research findings suggest, there is room for people's competence in this area to improve.

Students' Test Performances

We administered a test of observational ability to about 175 high school students. The test is based upon two stories, one a hiking trip and the other an automobile accident. Examinees are required to judge the believability of what characters in the stories claim to have observed. Statements are given in pairs, with the instruction to decide which, if either of the statements in each pair is more believable. The questions are based upon a set of principles for making such judgments. The principles, to be offered and discussed later, catalogue the effects which such factors as conflict of interest, degree of observational access, adequacy of technique employed and extent of independent corroboration, have on the trustworthiness of what people claim to have observed.

In school the students were given no explicit instruction in the set of principles and their use. Knowledge in this area would be an asset, though, since those with it could more effectively meet certain responsibilities such as, for example, serving as competent jurors. Our results indicate that there is considerable room for improving students' ability in this area, but some things appear to be known by the majority of students. Scores on the test averaged 49%, and ranged from less than 2% to a high of 74% ninety per cent of the scores were less than 65%. These results were obtained on a test which had about a sixth grade reading level, which in general seemed to capture students' interest and diligence, and which seemed to be well understood by most. The performances are thus plausibly explained by students' ability in the area. This explanation recei-

ves additional support from interviews, described in the following section, in which students described their thinking as they responded to the question on the test. We judge this performance to be indicative of inadequate ability to appraise observations.

The number of students who seemed aware of effects of particular factors on the reliability of observation reports varied greatly from factor to factor. There were areas in which the group as a whole performed well. They tended to perform best on questions requiring them to judge an observer's report on the basis of:

- (i) the observer's degree of skill;
- (ii) the observer's understanding of the situation;
- (iii) the precision and working condition of any instrumentation used;
- (iv) the degree of conflict with other dependable sources of information;
- (v) the reputation of the source of the information provided;
- (vi) the observer's conflict of interest; and
- (vii) whether the person was reporting on something previously forgotten.

About one-half to three-quarters of the students did well on these questions, an encouraging result in our opinion.

Students performed poorest when required to judge the effects which the following have on the believability of reports of observations:

- (i) responding to leading questions;
- (ii) reporting on non-salient features of an event, or state of affairs;
- (iii) being exposed to postevent experiences, such as police interrogation;
- (iv) reporting an observation in an environment different from the one in which it was made;
- (v) reporting an observation which someone else has made;
- (vi) reporting observations more precisely than appropriate; and
- (vii) the number of things which the report commits the speaker to holding true.

About one-third of the students were able to answer correctly questions based upon these effects. In addition, only about one-third recognized that observation reports tend to be more believable than inferences based upon them. Note that on a three-choice test such as the one used, one-third correct is the level of performance one would expect from purely random guessing. For principles not mentioned, students tended to perform between the one-third and three-quarters correct levels. The general tendency was for about one-half of the student's to answer questions correctly.

Taken as a whole we believe these results indicate too low a level of competence in this area of informal logic and corroborate the results of eyewitness testimony research. Only about one-fourth of all the principles were known by at least two-thirds of the students. There were pockets of strength, which is encouraging, but overall much improvement is desirable.

Thinking Processes

As one aspect of validating our observational ability test, we attempted to determine whether good performance on the test was the result of good thinking and whether poor performance was due to poor thinking. Since the test is in multiple-choice format, people's thinking is not apparent from their answer sheets. To uncover this thinking we interviewed a sample of about one hundred students while each answered about one-fourth of the items on the test. We initially instructed them to report all they could about the thinking which led them to choose the answers they did. This open-ended directive was sometimes supplemented by more leading questions, depending upon the completeness of response it elicited. Besides finding that indeed good and poor performance tended by and large to be related, respectively, to good and poor thinking, the interviews revealed much poor reasoning, some of which is described below. We achieved more than we had planned. In addition to information on the quality of our test we were able to produce a first-stage classification of the types of thinking errors people make when appraising observations.

As described previously, each question on the test presents two statements, and test takers are instructed to choose which, if either, of the statements they have more reason to believe at the time the statements are made. We found in many cases that students were unable to make a choice between statements, even when grounds for choice were present. Often, they asserted that it just seemed to be two people's opinions or that either statement could be right or wrong. We attributed this failure to see relevant differences between statements to lack of a broad knowledge of the types of factors which affect believability. In addition we found many situations in which students would act upon a condition such as an observer's expertise, but would not recognize the influence of other factors, such as being in a conflict of interest, or being emotionally upset. This indicates, confirming the performance results described earlier, that students are aware of the effects of some factors but not others, and suggests that they might profit from a systematic and comprehensive treatment of the types of factors which affect believability.

The most prevalent forms of poor reasoning were the following:

- (i) accepting or rejecting one of the statements at face value;
- (ii) giving a counterexample to disprove one statement, when it or a similar counterexample could readily be applied against the other;
- (iii) giving an example to support one of the statements, when examples to support the other are readily available;
- (iv) relying on an expert in a certain field regardless of the area in which the expert is asserting something or relying only on statements of experts;
- (v) accepting a statement on the grounds that it is more definite (lacks qualifiers such as 'generally', 'often', 'maybe', etc.) or that it is made by a person who seems more confident.

Other types of poor reasoning appeared but we will not mention all of them here. Rather, we will describe a couple of examples to illustrate the types of poor reasoning already mentioned. The examples from which these were chosen provide strong evidence of poor thinking in this field, and also point to places where remediation might be beneficial.

The first example is of the fourth type of poor reasoning. At about the halfway point in the test Student 32 had already recognized that in some questions a decision can be reached because one of the statements is made by an expert in the field. However, this thinking principle becomes distorted as the student begins to make choices for one statement or the other **only** when an expert is involved. This is illustrated in Student 32's response to the following question. In it two people disagree on the type of bird that was just spotted. Examinees are told that one of the people got only a quick glimpse. The keyed response is that the other person's statement is more believable because of this. Student 32 does not recognize this factor, however.

Question 32:

Meanwhile, Mary, Juanita, Scott, and Cheng are walking through the campsite. Scott points to his right and says, "**Look, there are two Swallows.**" Mary, who had been looking to Scott's left, turns quickly in the direction Scott pointed. She gets a quick glimpse of the birds and says, "those birds are not swallows. **they are Chickadees.**"

Student 32:

"I choose neither because neither one of them are experienced in the field. Like, they don't know what it is. If they had to be Professor Plant [**a biologist**], well, he knows and studies like things. So he'd know more than them two because they are not experienced in the field."

In the second example, Student 29 chooses incorrectly on the basis of the degree of confidence shown by the person who made it, and is typical of what we mean by the fifth type of poor reasoning. This error also appears prominently in eyewitness testimony research. In the question, some people are disputing the type of bird which had just been spotted.

Question 33:

Juanita says, "I was looking in the same direction as Scott. I saw the birds, too. **They were Sparrows.**" Scott becomes upset at what Juanita says. He shouts, "I know what I saw. **The birds were Swallows.**"

It is our judgment that in the context of the story presented, Scott's becoming upset should tell against the believability of what he says. Student 29 thought otherwise, however, using Scott's insistence to **support** Scott's statement.

Student 29:

"(I choose) the second. It says he shouts. He seems a bit confident that he knows what he saw, because he is expressing his point a lot better than the other one. She just said the birds were chickadees but he's trying to express his point stronger. He's shouting."

It is true indeed that many arguments are often **won** because one person is more forceful than the other, but this is not the ground on which they should be decided.

But for space, many other types of errors of thinking in judging observation statements could be discussed and illustrated. Suffice it to say that our experience with high school students indicates that they do not think well enough in this area. The eyewitness testimony research indicates that as adults they are also likely to not think well enough. The problem stems in large measure, we believe, from the lack of attention in the traditional curriculum to teaching techniques for appraising properly what others say, despite a need for these techniques in everyday affairs. In the following section we offer some suggestions for how thinking in this area might be fostered.

A SKETCH OF HOW TO ORGANIZE TEACHING

The evaluation of reports of observations is without doubt a complex activity requiring a sophisticated ability. Teaching such involved things to people is facilitated if the subject matter can be presented systematically and in small units. This approach has a tradition of some success in subjects such as logic and mathematics. Towards this end we offer in the table which follows a set of principles, which is a

modification of other work in this area (Ennis, 1980; Norris, 1979) relating the believability of reports of observations to several factors. Each principle is a tendency statement. Many of them are in

the form of empirical generalizations, and are subject to the same sort of evaluation as any claim of this type. Principle I compares the believability of observation statements to inferences based upon

PRINCIPLES FOR APPRAISING OBSERVATIONS

- I. Observation statements tend to be more believable than inferences based upon them.
 - b. has a suitable range of application;
 - c. is good quality;
 - d. works in a way that is well understood;
 - e. is in good working condition.)

- II. An observation statement tends to be believable to the extent that the **observer**:
 1. Is functioning at a moderate level of emotional arousal;
 2. is alert to the situation and gives his or her statement careful consideration;
 3. has no conflict of interest;
 4. is skilled at observing the sort of thing observed;
 5. has a theoretical understanding of the thing observed;
 6. has senses that function normally;
 7. has a reputation for being honest and correct;
 8. uses as precise a technique as is appropriate;
 9. is skilled in the technique being used;
 10. has no preconceived notions about the way the observation will turn out;
 11. was not exposed, after the event, to further information relevant to describing it; (If the observer was exposed to such information, the statement is believable to the extent that the exposure took place close to the time of the event described.)
 12. is mature.

- III. An observation statement tends to be believable to the extent that the **observation conditions**:
 1. provide a satisfactory medium of observation;
 2. provide sufficient time for observation;
 3. provide more than one opportunity to observe;
 4. provide adequate instrumentation, if instrumentation is used. (If instrumentation is used in gaining access, then the statement tends to be believable to the extent that the instrumentation:
 - a. has a suitable precision;

- IV. An observation statement tends to be believable to the extent that the **observation statement**:
 1. commits the speaker to holding a small number of things to be true;
 2. is corroborated;
 3. is no more precise than can be justified by the observation technique being used;
 4. is made close to the time of observing;
 5. is made by the person who did the observing;
 6. is strongly believed to be corroboratable by the person making it;
 7. does not conflict with other statements for which good reasons can be given;
 8. is made in the same environment as the one in which the observation was made;
 9. is not about an emotionally-loaded event;
 10. is the first report of the event provided by the speaker;
 11. is not given in response to a leading question;
 12. does not report a recollection of something previously forgotten;
 13. reports on salient features of an event; (Features of an event are salient to the extent that they are extraordinary, colourful, novel, unusual, and interesting, and not salient to the extent that they are routine, commonplace and insignificant.)
 14. is based upon a reliable record, if it is based upon a record. (If an observation statement is based upon a record, then the statement tends to be believable to the extent that the record:
 - a. was made close to the time of observing;
 - b. was made by the person who did the observing;
 - c. comes from a source having a good reputation for making correct records.)

them. Principles II, III, and IV relate believability respectively to characteristics of the observer, the observation conditions, and the observation statement itself. Each of these latter three principles consists of several subprinciples. They are presented in the columns of the table. The principles provide the units, or bite-sized portions, around which a curriculum might be constructed, though we will be unavoidably vague in this paper about actual classroom practices which might be used.

The principles must be interpreted cautiously. It is not proper to treat them severally as either necessary or sufficient conditions for observation statements to be believable. Rather, they must be applied judiciously, taking into account the characteristics of the situation at hand and relevant background knowledge, including experience in related matters. Application of the principles to actual cases is not a trivial matter, and comprises a competence over and above knowing and understanding the principles themselves. Thus, instruction cannot properly proceed without attention to the making of practical decisions.

Merely offering this set of principles as a teaching aid, together with the above brief remarks on their use, does not of course answer all important pedagogical questions. The following problems, among others, remain.

- (i) How are the principles and their application best taught? Can teaching be accomplished by presenting the principles all at once as a set, and illustrating them with a few examples? Must they be taught individually, always in the context of particular applications and discussions of their scope and meaning?
- (ii) How is knowledge of the principles best evaluated? Can evaluation effectively be done using traditional objective style tests, an approach we are now using? Must evaluation be carried out in more uncontrolled situations, similar to those in which the principles will eventually be applied?
- (iii) How are the appropriate principles accessed on particular occasions? Do people need to scan the entire list each time in order to find the principles which apply? Can the appropriate ones be accessed randomly, memory having been triggered by cues from the context being examined?
- (iv) Do the principles form a complete set required for evaluating observation reports?

These questions must remain unexplored here. We briefly look, instead, at another fundamental issue, namely the justification of the principles as sound inference rules.

JUSTIFYING THE PRINCIPLES

Nelson Goodman once argued (1965) that inferential principles and particular inferences are mutually justified by being brought into agreement with one another. He claimed: "A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend" (1965, p. 64). As recognized by Stephen Stich and Richard Nisbett (1980), the 'we' in Goodman's directive is bothersome. If the 'we' refers to the adult population at large, then, as Stich and Nisbett have demonstrated using results of psychological studies, many patently false inference rules and inferences would have to be sanctioned. These would include improper assignments of probabilities to events (such as believing the chances of flipping a head are increased after many tails are flipped successively), failing to account for statistical regression, and improper conclusions of causation from correlation data. In previous sections of this paper, we pointed out additional improper inferences which people commonly make. Interpreting 'we' broadly makes Goodman's rule dangerous, and thus was not applied in constructing our set of principles.

Stich and Nisbett modified Goodman's rule by replacing the 'we' with "the socially, consensually, designated authorities" (p. 201). Part of the justification for our principles is provided by Goodman's directive so modified. Many of the principles are based upon judicial practice. Rules of inference and of the admissibility of evidence have evolved in that field through the mutual adaptation of rules and particular inferences. Principles such as those dealing with, among other things, conflict of interest, observer expertise and reputation for veracity, quality of access to the things observed, leading questions, and record making, are all part of judicial practice. We judge this field to be an appropriate socially recognized authority in the assessment of reports of observations.

Judicial practice is not, however, the final court of appeal. At least two other sources of information are relevant. The first is psychological research into the factors which affect the accuracy of eyewitness testimony. This evidence is the basis of several of the principles including those about observing emotionally-loaded events, reporting on salient features, being exposed to relevant information after an event has occurred. In these aspects of assessment, psychological evidence provides grounds for altering as-

pects of judicial practice and inferential practice in general. Its role extends beyond merely describing the errors of reasoning which people make, into providing rather direct information on the way in which people **ought** to reason, a larger role than that seen by even some of the supporters of the relevance of psychology to the formation of rules of inference making (for example, Thagard, 1982).

Another technique for justifying the principles relies upon an approach similar to, and dependent upon inferring to the best explanation of an event. This approach might be called "inferring to the best prediction of an event". Inferring to the best prediction involves predicting an occurrence, and then pointing to a mechanism whose existence is plausible that would, if it were operative, produce that occurrence. Thus, for example, from the fact that a person was not alert while observing an event one can best predict that that person's accuracy in reporting upon that event would be diminished. This prediction is supported by a plausible accuracy-reducing mechanism involving the person's not attending the event, and thus of memory traces of the event not being stored, or not being stored as accurately, in the person's memory. Upon inspection, many of the principles receive support through such an approach, although, the details cannot be outlined.

SUMMARY

This paper provides a brief outline of several issues in an important aspect of practical reasoning: assessing reports of what people observe. It was demonstrated first of all that people's ability in this area is inadequate. Although strong in some respects, it is weak overall. Our evidence included test performances of high school students on a test of observational ability, reports of students' thinking as they chose answers on the test, and results of studies into the evaluation of eyewitness testimony. To help remedy this situation we have offered a set of principles for judging the believability of reports of observations and have given some brief indication of how they might guide a teaching effort. Finally, we presented a brief argument in defense of the principles as justified rules of inference. The argument rested on the traditional use of many of the principles by the judiciary, on support from eyewitness testimony research, and on an appeal to common experience about the likely effects of many of the factors on the accuracy of what people report.

REFERENCES

Brown, E., Deffenbacher, K., and Sturgill, W. "Memory for faces and the circumstances of encounter." *Journal of Applied Psychology*, 1977, **62**, 311-318.

- Ennis, R.H. "A conception of rational thinking." In J.R. Coombs (Ed.), *Philosophy of education* 1979, Normal, Illinois, The Philosophy of Education Society, 1980.
- Goodman, W. *Fact, Fiction, and Forecast*. Second Edition. Indianapolis: Bobbs-Merrill, 1965.
- Hagen, J.W., Meacham, J.A., and Mesibov, G. "Verbal labelling, rehearsal, and short term memory." *Cognitive Psychology*, 1970, **1**, 47-58.
- Leippe, M.R., Wells, G.L., and Ostrom, T.M. "Crime seriousness as a determinant of accuracy in eyewitness identification." *Journal of Applied Psychology*, 1978, **63**, 345-351.
- Lindsay, R.C.L., Wells, G.L., and Rumpel, C.M. "Can people detect eyewitness-identification accuracy within and across situations?" *Journal of Applied Psychology*, 1981, **66**, 79-89.
- Loftus, E.F. *Eyewitness testimony*. Cambridge, Mass.: Harvard University Press, 1979.
- Norris, S.P. *The dependability of observation statements*. (Rational Thinking Reports, Number 4). Urbana, Illinois: The Illinois Rational Thinking Project, 1979. (ERIC Document Reproduction Service No. ED 183 590).
- Stich, S.P., and Nisbett, R.E. "Justification and the psychology of human reasoning." *Philosophy of Science*, 1980, **47**, 188-202.
- Thagard, P. "From the descriptive to the normative in psychology and logic." *Philosophy of Science*, 1982, **49**, 24-42.
- Wells, G.L., Ferguson, T.J., and Lindsay, R.C.L. "The tractability of eyewitness confidence and its implications for triers of fact." *Journal of Applied Psychology*, 1981, **66**, 688-696.
- Wells, G.L., Leippe, M.R., and Ostrom, T.M. "Guidelines for empirically assessing the fairness of a lineup." *Law and Human Behaviour*, 1979, **3**, 285-293.
- Yarmey, A.D. *The Psychology of Eyewitness testimony*. New York: The Free Press, 1979.

NOTES

Copies of the test, "Test on Appraising Observations", are available from Stephen Norris.

David Hitchcock made helpful comments on a version of this paper.

This paper is based on research funded by the Social Sciences and Humanities Research Council of Canada, Grant No. 418-81-0781. □

Dr. Stephen P. Norris, Institute for Educational Research and Development, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X8

Dr. Ruth King, Institute for Educational Research and Development, Memorial University of Newfoundland.