

view because it is "true," and he has some way of establishing this. Truth can be held with varying intensity, however. Kelly also claims, correctly I think, that I am making the Aristotelian distinction between theoretical argument that leads to truth and practical argument that leads to reasoned action.

I am less certain about his claim that evaluation persuades someone to act rather than persuades them that something is the case. Action is the ultimate goal of evaluation, but there are so many other considerations involved in action that it seems unlikely the evaluator would be able to assess, or even identify the major contingencies. It seems to me that evaluation persuades as to the worth of something. Under some circumstances this may be a course of action, but ordinarily the action entails additional considerations.

4. For an extended analysis of an evaluation as argumentation, see Appendix A. For an analysis of "naturalistic evaluation," see Appendix B.

## COHERENCE AND CREDIBILITY

### The Aesthetics

#### *The Drunken Driver*

Humankind lingers unregenerately in Plato's cave, still reveling, its age-old habit, in mere images of the truth.

Susan Sontag, *On Photography*, 1977: 1

Consider two different images of the drinking driver. One may imagine the ordinary social drinker who happens to overindulge, and who, missing a stop sign, is detained by the police, thereby getting into trouble. Or imagine the drunken driver, one who is habitually drunk, a reeling, stumbling, insensate hazard to everyone on the road, including himself. The image that one constructs of the driver who drinks has much to do with the recommendations for action that one might embrace as a means of curtailing drinking drivers.

AUTHOR'S NOTE: This chapter was published as "Coherence and Credibility: The Aesthetics of Evaluation," *Educational Evaluation and Policy Analysis*, Vol. 1, No. 5, September-October, 1979, pp. 5-17. Copyright 1979, American Educational Research Association, Washington, DC. Reprinted by permission. Helpful comments were provided by Gene Glass, Jo Friedman, Donald Hogben, Rochelle Mayer, John Nisbet, and Helen Simons.

Until the past decade, the drinking driver has been perceived as the social drinker, a civil problem susceptible to correction by legal actions such as imposed penalties and fines. Over the past ten years or so, however, a series of studies has been influential in changing the dominant image of the drinking driver so that he is now perceived as more of an habitual, pathological drunk. The resulting recommendations for remediation are medical and are directed at a small subset of offenders rather than at all drivers who drink.

The sociologist Gusfield (1976) analyzed the rhetoric of these studies—how data were presented to persuade the reader of the conclusions. What he found is provocative. It has been axiomatic in social science that although a novelist may persuade a reader with emotion, a scientist persuades only with logic. Gusfield referred to this idea of language neutrality as the “windowpane” theory of language: language will reveal reality transparently without embellishment. He demonstrated that in actual practice scientific studies are valued in part for their dramatic use of language.

Gusfield analyzed in detail the most influential of the drinking driver studies, “Identification of Problem-Drinking Among Drunken Drivers” (Waller, 1967). Using Burke’s (1945) categories of scene, act, agent, agency and purpose, Gusfield showed how Waller’s study works as a persuasive rhetorical device. Since the literary style of science is one of neutrality, the scientific study must appear not to be literary, not to be concocted from a personal point of view. In Waller’s drinking driver study, the neutrality of scene is set by the role of the author, a medical researcher, writing in the *Journal of the American Medical Association*. Once the neutral authority of the scene is established, the author as a person does not intrude.

The form of the report is narrative, with a definite dramatic structure. Tension arises over the choice between the two types of drinking drivers, the tension finally being released in a denouement. The dramatic progression is from the problem drinker as social drinker to the problem drinker as drunken driver. In the progression it is important, according to Gusfield, that the change in perspective (which is the actual outcome of the study) be seen as the culmination of external data derived by scientific method. To this end the voice of the report is impersonal: “Recent reports have suggested . . .”; “It is increasingly becoming apparent . . .” The passive voice places the action in the external agency of the data and method, rather than in the author. It reinforces the idea that the conclusions emerge from the impersonal data world.

The resulting “scientific” style is clinical, detached, impersonal, and lacks imagery. The author presents the external world and allows it to persuade the reader. The style suggests that the observer is governed by method and by the rules of scientific integrity. Attention to detail and meticulous description of procedures, such as carrying numbers to several decimal places, give the impression of accuracy. According to Gusfield, this style reinforces the basic epistemological assumption: by use of the same method, different observers must come to the same conclusions.

Although the discussion of the methodology is conducted with painstaking neutrality, the implications for action are handled quite differently. The facts are converted into imagery. In his study of drinking drivers, Waller repeatedly referred to “drunken drivers,” thus conjuring up powerful images, as opposed to describing in more neutral terms like “drivers who get into accidents after drinking.” Even the title of the study uses the term “drunken driver.” In actuality, the category of drunkenness was defined operationally by alcohol content of the blood, a legal definition carrying quite a different emotional loading than the image of the drunken driver.

According to Gusfield, once Waller identified “drunken drivers” as the problem drinkers through his survey of civil records, he drew the action implications from the imagery of his original classification. Thus the medical scientist made sense of his data by reduction to imagery. The drunken driver imagery extended the meaning of the primary data to meanings already commonly accepted and known by the audience. It was not a simple extrapolation of the data.

In substance, the drunken driver image attributes responsibility to the agent, to the driver himself, and not to the scene within which the driver acts. Drunken driving becomes an attribute of self which requires medical treatment. It implies pathological rather than normal behavior, and correspondingly different courses of remediation. Even a different subgroup of people is implicated. The civil delinquent is transformed into a patient.

Gusfield suggested that this study is typical in rhetorical style, although particularly influential, among the drinking driver studies. More generally, he argued that science works through such reduction. Generalizable knowledge is created through linking specific objects to universal categories. “It is implicit and inherent in the enterprise of defining, describing, and interpreting data through verbal or written communication insofar as conclusions and generalizations imply mean-

ings for action" (p. 31). And later, "To be relevant or significant, data must not only be selected, they have to be typified and interpreted" (ibid.).

Although not all evaluation studies employ such vivid imagery, similar issues arise in evaluation. Imagery, dramatic structure, and mode of presentation are central considerations for the import of an evaluation. These elements, often thought of as merely cosmetic, can affect what people believe and do. How do these elements function and what is their relationship to the content, the "truth value," of an evaluation? Evaluators do employ such elements, although not always consciously and not to the same degree. As Gusfield notes, "What is at stake, however, is the necessity of the interpretation and the close connection between that interpretation and its form of presentation, its artistic element" (p. 32).

Similarly, in his analysis of the role of images in defining problems in social policy, Schon (1979) contended that social problem-setting is mediated by the stories people tell about troublesome situations. The framing of the social problem depends on the metaphors underlying the stories. How the problems are framed is critical to the solutions that emerge. For example, a pervasive story about social services is that they are "fragmented," and the implicit solution is that they be "coordinated." But services seen as "fragmented" could also be seen more benignly as "autonomous." The underlying metaphor gives shape and direction to the problem solution.

Schon contended that we are guided in our thinking about social policy by pervasive, tacit images, which he calls "generative metaphors." They involve a carrying over of one frame of reference to another situation. Usually, these metaphors are induced by immersing oneself in the experience of the phenomenon. These guiding images are necessary to our thinking.

For example, there are two rather different views of urban renewal. One view sees the slum as a once healthy community that has become diseased. A social planner with such an image envisions wholesale redesign and reconstruction as the cure to urban blight. Quite a different view of the slum portrays it as a viable, low-income natural community which offers its residents important social benefits. The second view implies strikingly different prescriptions for improving the community.

The story of the slum as blight was dominant in social planning in the 1950s. In the 1960s the story of the slum as natural community

arose as a countermetaphor to vie for public and expert attention. According to Schon, from a reality that is "ambiguous and indeterminate," each story selects features that become the themes of what the story is about. In the first vision, terms like "blight," "health," "renewal," "cycle of decay," and "integrated plan" become important. In the second vision, "home," "patterns of interaction," "informal networks," and "dislocation" are key ideas.

Each story presents a view of social reality by selecting, naming, and relating elements within the chosen framework. According to Schon, "naming" and "framing" are the key processes. By selecting certain elements and coherently organizing them, those processes explain what is wrong in a particular situation and suggest a transformation. Data are converted to recommendations.

Naming and framing proceed by generative metaphor. The researcher sees the slum as blight or as natural community. In seeing *A* as *B*, the evaluation implicit in *B* is carried over to *A*. The first metaphor is that of disease and cure. The second is that of natural community (versus artificial community). The transferred evaluations are based on images deep-seated in our culture. Once we see a complex situation as health/disease or as nature/artifice, we know in which direction to move.

Seeing *A* as *B* greatly facilitates our ability to diagnose and prescribe. On the other hand, it may lead us to overlook other important features in the situation that the metaphor does not capture. Since generative metaphors are usually tacit, important features may pass undetected. Schon argued that we should be more aware of our generative metaphors, and that this is best done by analyzing the problem-setting stories that we tell.

The "deep" metaphor accounts for why some elements are included in the story while others are not, why some assumptions are taken to be true in spite of disconfirming evidence, and why some recommendations seem obvious. It is the image of the drunken driver or the metaphor of the slum as diseased that gives shape to the study and direction to one's actions.

### *The Story*

That is to say, the first thing we *do* with images is to envisage a story; just as the first thing we do with words is to



tell something; to make a statement. . . . Pictures and stories are the mind's stock-in-trade.

Susanne K. Langer, *Philosophy in a New Key*, 1942: 128

Consider an evaluation. First, there is a series of events, the reality as it were, which will always remain somewhat indeterminate. The evaluator faces an ambiguous world. Through various social and psychological screens, the evaluator portrays these events in a report. The report itself is an artifact. As such, the report has coherence and form, some kind of aesthetic structure, even if the structuring is not entirely conscious.

The report is interpreted by readers, and this interpretation will vary from reader to reader, depending on her circumstances and her background. If the readers find the report credible, they are more likely to assume a corresponding valuing position vis-à-vis the object being evaluated. Furthermore, the readers may be led to action, depending on their disposition and other circumstances. The action may or may not be recommended by the evaluator.

As an artifact the evaluation report will have aesthetic qualities, appearances that engage interest, but there is something more basic to an evaluation report than this. Every evaluation must have a minimum degree of coherence. The minimum coherence is that the evaluation tell a story. There must be either an explicit or tacit sequence of events (or more accurately, an interpretation of events) for the reader to use the evaluation as a guide to valuing. There also may be recommendations by the evaluator, but the recommendations are not necessary. The story is.

In this sense, images by themselves cannot be evaluations. As Sontag (1977) noted in her analysis of photography:

Desire has no history—at least, it is experienced in each instance as all foreground, immediacy. It is aroused by archetypes and is, in that sense, abstract. But moral feelings are embedded in history, whose personae are concrete, whose situations are always specific. Thus, almost the opposite rules hold true for the use of the photograph to awaken desire and to awaken conscience. The images that mobilize conscience are always linked to a given historical situation. The more general they are, the less likely they are to be effective [pp. 16-17].

Photographs may be used to reinforce valuations and even presented in sequence to tell a story (Templin, 1978). The story itself is neces-

sary, however, to interpret the events. Events must be presented in a specific historical context. The story, even more than the image and the metaphor, is the basic underlying structure of an evaluation.

There are at least two conventional ways of telling the story. One way is to present the evaluator as a neutral, scientific observer. In this case, the story line is implied. It runs something like, "I am a detached, neutral observer who has made measurements according to the canons of science and have found certain things to be so. The program was conducted as I have described it, and I have found the following outcomes. . . ." In establishing this voice, the evaluator meticulously specifies his methodological procedures to enhance his credibility. Readers are expected to believe the results because of the objectivity of the methodology and because of their previous experience and belief in such a methodology. Usually the story line concludes that "the program was implemented, and such and such were the results." Actual description is often sparse. The drunken driver study is an example of this story line. The usual presentation is to describe the project or the goals of the project, the treatment, the results or effects, and the conclusions.

The second major way of telling the story is for the evaluator to stand closer to the program, as reflected in the narrator's "voice," and to tell the story by describing the events in detail. To this end the evaluator may use emotionally charged language and a narrative presentation. The story may look like a newspaper report. The first approach to story telling lends itself to quantitative and the second to qualitative methodology. But in both cases there is an ordering of events that tell a story, even though many of the events are assumed rather than explicit. The more formal the presentation becomes, the more things are assumed.

One ought to distinguish the "story," which is an interpretive ordering of events basic to all evaluations and prerequisite to valuation by the reader, from dramatic form. The dramatic form may vary, and the same story may be presented in a number of different ways. An evaluative story may be made more compelling, interesting, and pleasing by the dramatic form and by other aesthetic elements. However, although the story is more basic, the aesthetic rendering is not merely cosmetic. Both good art and good science lead the reader to experience an event in somewhat the same way as the creator experienced it (Bronowski, 1956). As Gusfield showed in his analysis of the drunken driver studies, the aesthetic rendering can have an important effect on

the recommendations and meaning of a policy study, even a scientific one.

It is not too surprising that the "story" of an evaluation is more basic than the aesthetic rendering, since the story line relates events to each other in specific ways, such as in cause and effect relationships. The events of the story may be presented in different dramatic forms, but both the story and the aesthetic elements contribute to overall coherence. All things being equal, greater coherence leads to greater credibility on the part of the audiences. Things fit with more certainty. However, if the report is unrealistically coherent, credibility is lost.

One may think of the difference between the story and the dramatic presentation as the difference between "content" and "form," or in Polanyi's terms (Polanyi and Prosch, 1975), the difference between the "story" and the "frame." The frame may include imagery, meter, or other "artificial" modes of presentation. This is not to say that the "frame" is purely cosmetic. The meaning of a work is the integration of its appearance with its content. The meaning of a poem is not merely the content of the poem written in prose.

The story differs from a chronology of events in that the story consists of inferences and interpretations of events. Events are integrated with each other, and parts are ordered to the whole. The story itself can be more or less tightly integrated, and it provides the necessary coherence for the evaluation as a whole.

A key concept here is that of coherence. Coherence consists of logical connection, congruity, or "consistency in reasoning or relating, so that one part of the discourse does not destroy or contradict the rest; harmonious connection of the several parts so that the whole 'hangs together,'" as in the coherence of an argument or report (Oxford English Dictionary).

Following Polanyi (Polanyi and Prosch, 1975), I wish to expand the scope of "coherence" to include imagery, dramatic structures, and other aesthetic elements. For example, one might consider the logical consistency of the *concept* of the drunken driver as it is used in argumentation. But also, one might consider the *image* of the drunken driver as it is used harmoniously with other elements.

The image itself serves as a condensation of entangled meanings, emotional as well as cognitive, with its own dense coherence. It relates to the every-day, "real" world in a way that the concept does not. The concept can be defined explicitly, the image by a tacit sharing. The multilayered meanings implicit in images and other aesthetic elements

make it possible to convey complex, elaborate, and highly refined evaluations of the objects under study. The aesthetic elements convey intensity, unity, and complexity (Beardsley, 1958).

Aesthetic elements, like images and dramatic structures, are distinguished from logical entities in that aesthetic elements are apprehended immediately without recourse to formal arguments. In this sense they are like perceptions. Aesthetic elements inhere in appearance only, and they are apprehended by tacit influence, by an unconscious integration of their parts (Polanyi and Prosch, 1975). So if one reads the drunken driver study, one assimilates somewhat unconsciously the tacit meanings conveyed through the images and dramatic structure. There is a considerable strand of "subliminal perception" in the learning, though one can easily detect the aesthetic elements if they are pointed out.

Furthermore, the reader participates actively in such learning. The reader becomes actively engaged in integrating the parts into tacit meanings and often becomes deeply engaged by the process, though the depth of engagement will vary from reader to reader. Eventually, the tacit inferences and the explicit inferences will be integrated into one's global image, one's "subjective knowledge structure." It is one's world image that ultimately affects one's behavior (Boulding, 1956).

I would like to argue, explicitly not tacitly, that greater coherence makes the evaluation report more credible, more worthy of belief and confidence. An incoherent report is not credible at all, and the more coherent the report (up to a point), the more credible it will appear to be—all other things being equal.

Usually the evaluator also strives to be persuasive—to win the audience to a point of view or even a course of action through reason and understanding. Most of the burden of persuasion falls on explicit argumentation, but here too coherence reinforces the persuasiveness of the argument. Less frequently, the evaluator may even aspire to secure commitment, to bind or obligate the reader to a particular position. These are the most powerful of the evaluations, but they pay the price of reduced scope of audience in exchange for deeper commitment of a few. From credibility to persuasiveness to commitment, one moves closer to action. In most evaluations, coherence plays the humbler role of enhancing credibility.

### Form

Science is nothing else than the search to discover unity in the wild variety of nature—or more exactly, in the variety of our experience. Poetry, painting, the arts are the same search . . . for unity in variety. Each in its own way looks for likeness under the variety of human experience.

J. Bronowski, *Science and Human Values*, 1952: 16

If the content differs from the dramatic structure of its presentation, what is it that comprises aesthetic form? In a classic article on "The Problem of Esthetic Form," Parker (1960), following Aristotle, suggested that aesthetic form can be analyzed in terms of six principles: organic unity, theme, thematic variation, balance, hierarchy, and evolution. The master principle is that of organic unity, and, in different ways, the other principles are subservient to it. Organic unity enhances total coherence.

The principle of unity demands that the work contain only necessary elements. In a sense, aesthetic form is in contradiction to life, which contains many unnecessary elements from an aesthetic point of view. The other five principles are manifestations of the unity principle. In an evaluation report, one must have a "theme," and this theme must not only persist but reappear in variations if the total effect is to be achieved. Often in evaluation studies, this effect is achieved by presenting data which build a common theme. Such building and syncopation of data reinforce veracity. They mobilize the reader.

"Balance" entails an equality of opposing elements playing back and forth within the form. It provides dramatic tension. "Hierarchy" is one way the separate parts may be organized. Finally, Parker added the principle of "evolution," suggesting how a total meaning can be created by an accumulation in which earlier parts determine that which happens later. Evolutionary unity can be achieved by a climax at the end, or by the end itself assuming greater importance than the other parts. In an evaluation, the report may evolve into conclusions which have emerged from the data presentation.

These aesthetic principles are perhaps most easily seen in a dramatic narrative. Woodward and Bernstein's *The Final Days* (1976) is not an evaluation, but it is an interesting treatment of presumably factual material. The final days of the Nixon regime are recounted day by day. This daily countdown to the inevitable denouement when Nixon resigns

adds to the dramatic intensity of the piece. Woodward and Bernstein do not simply relate one person's account of events and then relate another's, which must have been the way the information was collected. Rather, they blend various perceptions into an overall complex pattern, shifting from the thoughts and actions of one person to those of another. From this interweaving, certain themes emerge.

Forces converge on a defiant and recalcitrant Nixon. One by one, in differing ways, his closest supporters become convinced that he is guilty and must resign. These little personal vignettes within the larger drama are repeated in thematic variations which contribute to the overall theme of Nixon's support continually eroding. Nixon himself is uncertain as to whether to resign or to stay in office. This balance shifts day by day, hour by hour. Growing from the theme of Nixon's increasing isolation emerges a transcendent theme—his personal disintegration. These themes are skillfully woven together until they evolve into the denouement of his disgrace, his resignation from office, and his collapse as a man.

Similar aesthetic principles apply to quantitative material, though their effects are sometimes less easily discerned. The simplicity, economy, and elegance of a particularly appropriate experimental design or multivariate analysis are manifestations of the unity principle. In fact, some scientist-philosophers like Polanyi believe that in the higher reaches of thought, the mathematician is guided primarily by his aesthetic intuition.

Consider, for example, the aesthetic appeal of a multivariate analysis as compared to a long string of bivariate data analyses. Such compression harbors an appeal beyond the fuller truth value a multivariate analysis might entail. It is at least partially in such satisfactions that the pursuit of more elegant designs lies. Or consider the nested appeal of a Guttman scale. The parsimony of such a scale, and of hierarchical theories based on such an ordering, appeals to aesthetic sensibilities. One even looks for problems and data that will fit such designs.

The aesthetic appeal of quantitative studies is reflected in terms such as "coherence," "elegance," "economy," and "power." There is a sense in which "elegant" solutions are compact, simple, swift. They give one a sense of control, utility, and mastery similar to searching for and finding the "right word" in describing something. Elegant solutions dispense with unnecessary steps and procedures. They are compelling.

It is not only in quantitative studies that one finds such coherence. Nor must the structured elements be numbers or symbols. One of the



primary reasons for the persuasive nature of Glass's (1972) evaluation of AERA instructional tapes is the way he fits the arguments together into an overall pattern (Appendix A). The arguments are the elements. Glass achieves an integration of the logical parts of the evaluation within a coherent overall rhetorical structure. Elegance in this case means that there are few parts not necessary to the overall logic of the evaluation. There is a clear, if complex, guiding structure with the parts fitted neatly and properly to the whole.

In addition to elegance and coherence, one might say of Glass's work that it is credible because of "voice," the attitude of the narrator toward his subject matter. Glass projects a "persona" such that it appears the narrator will be convinced by the data no matter how the data turn out. The image of the narrator is that of a "rational" man ready to be persuaded by the evidence. To identify this as an effect of style is not to question its authenticity.

### *"Outward Bound"*

To evoke in oneself a feeling one has once experienced and having evoked it in oneself then by means of movements, lines, colors, sounds, or forms expressed in words, so to transmit that feeling that others experience the same feeling—this is the activity of art.

Tolstoy, "The Communication of Emotion"

Smith, Gabriel, Schott, and Padia's (1976) evaluation of Outward Bound was unusual in that it used both quantitative and qualitative methods. The two parts were handled separately, and both were included in the final report. The opening lines of the study set the problem and dramatic tension:

What happens to the young person who elects to undergo the physical challenges associated with the rocks, cliffs, snowfields and streams of the Rocky Mountains and the interpersonal challenge of a small group isolated in a foreign environment? Proponents of the Outward Bound program, alumni and staff, maintain that the psychological effects are unique to each participant and cannot be anticipated. In contrast, many others claim that the program causes profound, predictable changes in the participant's feelings about himself and others [p. 401].

The tension introduced in these opening lines provides drama, significance, and interest. The basic question was addressed in two ways—by quantitative measurement and by the qualitative participant observation. Both reflected the overall evaluation story.

In the quantitative part four variables were chosen as outcome criteria: "self-esteem," "self-awareness," "self-assertion," and "acceptance by others." Psychometric instruments were developed to measure all four variables. The "voice" of the quantitative section was that of the neutral, scientific narrator. An opportunity for elegance came in the design of the study.

With groups of participants proceeding through the training at different times during the summer, the evaluators randomly assigned students to groups and administered the questionnaires only once to each group at different times. Thus they were able to employ a complex time-series design in which groups acted as controls on each other. Forty-four means were collected and analyzed for each of the June, July, and August courses.

Unity was achieved by neatly graphing the forty-four means so they could be compared to each other visually as well as mathematically. A four-by-three table succinctly summarized the considerable mass of data. The only aesthetic failing was that the four outcome variables were not related to each other in any way, thus disrupting the unity of the analysis. The same basic question was also addressed by participant observation. Dramatic tension was established by a first-hand account. The style was personal and involved, rather than neutral and detached.

The six-hour bus ride from Denver to the San Juan Mountains let our fears and expectations incubate. My own fears centered on the difference between myself and my fellow Outward Bounders. I guessed that I was twice as old as the rest and wondered if their strength and endurance would surpass mine. The prospect of failing physically and being a social outcast was cheerless. Like the children around me, I knew something of what Outward Bound was supposed to be and to do to me. I was aware of its military origins and the rumored physical dangers of the program. There was also the intimation that one would experience the mystical entry into adulthood. Although I had long since entered adulthood, certain aspects of myself raised doubts about the coming 23 days; my fear of heights, my inexperience with competitive situations, my intransigence when pushed. The doubts were repeated mental questions. Would I be left behind? Would I panic

halfway up some steep rock face? Would I fall into an adult role, take too much responsibility, and be rejected? Would I quit? The silence on the bus led me to believe that the others were also ruminating on the experiences ahead. We each had our own private visions and nightmares [pp. 411-412].

Interest in the study was intensified by the personal drama and by the expression of feelings. The headings of the separate sections indicated the dramatic progression: "Expectations," "Competencies," "Expeditions," "Solo," "Social Challenge," "Final Expedition," "Marathon." The climax of the narrative was a dangerous mountain ascent and an exhausting race. The writing was both insightful and colorful.

I decided that Outward Bound was especially designed for people like Chris. She had never worked, never been challenged, never really lived except vicariously through her parents and the media. The course forced her into a series of compacted experiences, causing her maturation to be accelerated. One could almost see her stamina and confidence grow as the course progressed. She tried hard, but never stopped complaining. She could not give up her play of professing weakness and hiding her strengths. The strategem must have worked for her; didn't it get her a free ride across the Rio Grande on my back? [p. 417]

The neutral voice was abandoned for an intense first-person account of events. Outcomes like self-assertion, self-esteem, and self-awareness were registered in personal manifestations. This rendering of events provided a different level of understanding and meaning than the findings that the Outward Bound course had a positive impact on participants' "self-assertion" and "self-esteem" but did not affect measurably participants' "acceptance of others" or "self-awareness," which were the results of the measurement of the four variables in the quantitative part of the evaluation. However, both sections were structured aesthetically. Together, the two components provided a highly qualified but credible and persuasive answer to the question set in the first paragraph of the study.

Together, the two analyses tell only one story about what happened, and the same story could have been presented in different dramatic forms. For example, the overall structure might have been a dramatic narrative throughout, with the quantitative data spliced in for support.

Or the report might have consisted primarily of the quantitative analysis with a bare-bones explanation of the effects included in the "discussion" section. Alternatively, the evaluation report might have been written as a Socratic dialogue or a play. Smith et al. (1976) chose to give equal weight to the quantitative and qualitative.

Whatever the form, the evaluation story would be the same, though it might differ in aesthetic power. The credibility of the evaluation depends partly on the correspondence of the quantitative part to the principles of measurement, the correspondence of the qualitative part to the personal experience of the audience, and the correspondence of the parts to each other—their coherence. If all these elements fit together properly, and the evaluation is aesthetically rendered, then the entire evaluation will be seen as highly credible. In this case, the parts do cohere, and the audience is likely to be convinced of the efficacy of the Outward Bound training. The one failing is that the quantitative and qualitative parts are not explicitly united, so the overall unity is not as intense as it might have been.

Whether an evaluation abounds in metaphors or coefficients, dramatic structure or mathematical economy, every evaluation must do one basic thing: it must tell a story. The story may be explicit, truncated, or implicit, but the story must be there if one is to draw implications. An evaluation must interpret events occurring across time.

In many quantitative evaluations the story is often assumed. For example, in the Outward Bound quantitative evaluation, one assumes that there was a program of some kind, a series of events, and that the evaluation instruments measured the outcomes of these. The purpose of the experimental design was to find out if the assumed events were the causes for the changes in the measures. If so, certain inferences follow. Whether the evaluation was aesthetically pleasing resides in the elegant use of the experimental design, the statistics, and the like.

In the qualitative section of the evaluation, events were portrayed in a personal way, rather than assumed. The aesthetics resided in the images and dramatic structure. Direct observation in every-day categories rather than readings of scientific categories on instruments were the sources of data. The inferred story mobilizes the reader, and the better the presentation of the story is aesthetically, other things being equal, the more the reader is mobilized.



### Authenticity

A work of art is authentic or true not by virtue of its content nor by its "pure" form, but by the content having become form.

Herbert Marcuse, *The Aesthetic Dimension*, 1978: 8

Why is it that aesthetic elements like images and dramatic structure carry such import? There are many explanations, mostly drawn from theories of aesthetics or art, but no single accepted explanation (Dickie, 1971). It is clear that we see our lives in the form of scenarios and stories, that images are shaped in the language of our every-day thoughts. Concrete images are the stuff of which our minds and memories are composed.

The anthropologist Turner (1973) contended that affect is "inherently concrete, particular, and associated with the unique relationship of the self to its objective environment" (p. 354). Abstract principles are inadequate for acting in the real world. Such an integration into the environment must be accomplished through particular, affective condensations such as only imagery, symbol, rituals, and myths can supply. These devices convey "affective and motivational power" within the subjective meaning framework of the person.

Similarly, Sullivan (1977) suggested that it is the role of the imagination to provide integrating images and myths. The individual always acts within a particular personal and historical context. Highly formal and abstract theories of action and behavior overlook this deeper level of contingent motivation.<sup>1</sup>

While such analyses are astute, they may underestimate the mythic power of science itself. People do sometimes act on the basis of results they consider to be "scientifically" derived. Science has both its own myths and its own mythic authority. Nor does the necessity of concrete imagery entirely explain the effects of the more abstract aesthetic patterns.

For example, why is dramatic form so effective? According to Broudy (1972), although dramatic form is not equivalent to moral or metaphysical meaning, it is the first "intimation" of it. Literal truth is usually confusing and insignificant. It must be converted into "plausible fictions." Illusions are necessary for human import. Aesthetic renderings become searches for significance in human events. Furthermore,

people often judge the credibility of reported events on the basis of aesthetic criteria like vividness and inherent unity—on appearances.

An aesthetic experience, then, is *an* experience rather than merely experience. The artist selects events that are cumulative in impact. Tension arising from the dramatic structure commands the audience's attention. For example, in the drinking driver study, the researcher maintains tension between the two competing images of the drinking driver. In Broudy's view, the alternative to imaginative portrayal is insignificance. Aesthetic presentation tries to capture the "essence" of events rather than their literal reality.

Polanyi, too, sees significance as the underlying issue:

Our lives are formless, submerged within a hundred cross-currents. The arts are imaginative representations, hewn into artificial patterns; and these patterns, when jointly integrated with an important constant, produce a meaning of distinctive quality. These artificial patterns are . . . what isolate works of art from the shapeless flow of personal experience and public life. They make of works of art something detached, in many cases portable and reproducible, and potentially deathless [p. 101].

From a more radical perspective, Marcuse (1978) emphasized the importance of both personal feeling and artistic form in social reform. Subjectivity—the inner, personal history of the individual—can be a liberating force when it is expressed in a work of art. The "given reality," the reality shaped and verified by the dominant social institutions, can be reshaped in the art work. The art work constitutes an alternative reality which is in opposition to the "given" reality. Thus an art work challenges the monopoly of the established institutions in defining reality.

In this process, aesthetic form transforms the content into a new, self-contained whole. This new vision reveals the repressed dimensions of reality in a "fictitious reality" which captures only the essence of events. Through concentration, exaggeration, emphasis on the essential, and a reordering of facts, the audience's consciousness is restructured. Experience is intensified.

The content of the work is subjective feeling, and aesthetic form transforms the content into a self-contained whole, enabling the work to stand against the "given reality." When the content has been properly converted into form, the work has the ring of authenticity. Thus Marcuse can say, "The authentic utopia is grounded in recollection" (p. 73).

*The Beauty of the Drunken Driver*

Wisdom, the ability to go to the heart of the matter in concrete situations, is acquired slowly; it is a discipline of experience, imagination and story, not of naked intelligence.

M. Novak, *Ascent of the Mountain,  
Flight of the Dove*, 1971: 58

The documentary film maker Frederick Wiseman, well known for his films portraying social institutions, such as the welfare system and the meat-packing industry, described his television documentaries as "reality fiction," as a "fictional structure with the illusion of truth" (O'Connor, 1976). From the "formless glop" of fifty hours of film, Wiseman shapes the documentary in the editing process, where he introduces the elements of pacing, style, and structure. Although selecting arbitrarily, the film-maker attempts to capture a portion of the "truth" as he sees it, according to Wiseman.

Yet evaluation is not fiction, or at least it should not be. The dicta that apply to art, even documentary art, are not necessarily those that shape evaluation. One expects more veracity from an evaluation report than from a novel or a film. To return to the original question, how can one justify artistic presentation of data that affects the very interpretation of the study itself?

Langer (1942) suggested that there are at least two major types of meaning, the literal and the artistic. Literal meaning deals with explicit content and is expressed as propositional fact. It can be either true or false. Artistic meaning deals with tacit form and is expressed as feeling. It can be either adequate or inadequate. Correspondingly, factual reference and poetic significance are two rather different relations between symbols and their meanings, but both can be identified.

In Langer's theory, science is advanced by discursive thought which is reducible to analytic concepts and to "facts." Langer points out that the discursive use of symbols, which reaches its climax in science, is only one way of knowing. There are nondiscursive ways of knowing, like visual forms, forms in which the elements of the whole are presented simultaneously as in paintings, images, metaphors, and myths. As such, artistic meaning is not subject to tests of falsification in the same way as are propositional statements.

Rather, artistic form is more comprehensible in terms of consistency, coherence, economy, and elegance. However useful the discursive symbols, one must turn eventually to artistic form for the expression of feeling. Langer's theory of art is an "imitation" theory in which artistic form serves as analog to emotion, just as discursive language serves as logical analog to the physical world (Dickie, 1971). Whatever the strength of the claim, it is undeniable that artistic meaning can sometimes conflict with literal meaning. It is entirely possible that inappropriate uses of metaphor or wrongly reconstructed events or omitted data can mislead an audience.

For example, Woodward and Bernstein's *The Final Days* evoked questions about its truthfulness. A controversy developed over whether the portrayed events were true and even to what extent the authors could know certain private happenings. Some participants contended that the authors had left out discordant details in order to intensify the drama about Nixon's disintegration. The credibility of the authors suffered because the book was too dramatic. Whatever the merits in this particular case, it is not difficult to see how literal truth may sometimes be sacrificed to dramatic form. John Keats notwithstanding, truth and beauty are sometimes incompatible.

Beyond the truth of the work itself is the question of whether the work is fair to the people portrayed. I have dealt with this justice issue in another place by connecting it to infringement of one's right to self-esteem (House, 1976). The trade-off between truth and justice in *The Final Days* is certainly debatable. Even liberal newspaper columnists have protested that the authors went too far in their portrayal of Nixon as a disintegrating person.

An evaluation can be seriously threatened if an element in it proves to be untrue or of questionable origin. What if, for example, in the Smith et al. (1976) evaluation of *Outward Bound*, one discovered that the person who wrote the excellent first-person narrative was not the person who actually experienced the travail? Somehow it would threaten the authenticity of the piece if the writer reconstructed someone else's adventures.

In fact, what if the trip into the Rockies were really based on three trips taken at different times by three different persons, and for aesthetic reasons, the author combined the three separate experiences into a single dramatic narrative? The advantage would be obvious. As it stands, the narrative powerfully conveys the feeling tone of the mountain experience—truth of feeling—in a way that three fragmented narra-

tives could not do. The advantage would be a gain in the coherence of the piece and in its credibility and persuasiveness.

On the other hand, the narrative would not be literally true. Would the author be justified in taking such liberties? There would be a trade-off between literal truth and truth of feeling, between truth and beauty. As important as beauty is in evaluation, truth must take precedence when there is a conflict between the two. Ultimately, evaluation is not fiction, as necessary as imagery and form may be to conveying complex meaning. Trade-offs are possible between truth and beauty. One may omit some facts to make the report more coherent, more readable. But one would sacrifice only a little truth for a great deal of beauty. Truth is relatively more valuable.

For me the resolution revolves around the idea of the evaluative story. Liberties are permissible with the dramatic structure as long as the evaluative story remains unchanged. That is, did people go into the mountains and experience these things, although not exactly as the literal plot would have it? If people did not experience similar things or if the figurative language suggested feeling states that were different from those the mountain trip induced, then the license with the evaluation would be too great. The audience would be misled in a significant way and have a false picture of the program. Also, one would need to be told that the events were reconstructed. One could test the veracity by the internal consistency of the narrative, by its coherence with the quantitative analysis, and by its congruence with the reader's own experience.

Nor are quantitative studies immune from this conflict. Measurement and analysis techniques are often based on aesthetic criteria and sometimes result in inappropriate applications. Furthermore, the methodology sections in most studies are written from a reconstructed view of what happened rather than how the data were actually collected and analyzed. There should be a limit to such reconstruction, the limit being the point at which the evaluation story is significantly changed. For example, the audience may assume that standard precautions in data collection were taken when indeed they were not. Or the audience may assume a treatment occurred when, in fact, its implementation was questionable. Since in a quantitative study so much is assumed by the audience, and unspoken by the authors, reconstruction of what happened is a continuing problem.

To return to the initial example of the drunken driver study, was the researcher justified in employing potent images of the drunken driver?

The answer is a qualified "yes," if one accepts that the purpose of an evaluation or a policy study is to inform opinion and to lead to action. The qualification is that the image of the drunken driver must be congruent with the rest of the study, though not the only rendering the author might have chosen, and that the rendering must reflect properly the real world. Such renderings are essential to moral action. They speak the "vocabulary of action" that remains an obligation of the evaluator (House, 1973). Yet in those concrete instances in which truth and beauty conflict, truth is more important than beauty. And justice more important than either.

## NOTE

1. Similar theories exist in aesthetics. The objectivism of Piaget's psychology has its analogs in both moral philosophy and aesthetics. In philosophy Rawls's (1971) theory of justice is predicated on an "original position" in which a "veil of ignorance" limits the knowledge and motivations that people choosing their principles of justice must have. People in the hypothetical original positions do not know their social position or particular talents. In addition, they are rational and not altruistic. In other words, self-interest has been reduced as motivation, and people presumably will act impartially. As in Piaget's psychology, the actor is conceived as an impersonal, abstract being without contingent motivation.

In the philosophy of aesthetics the corresponding position is called phenomenological objectivism (Gottshalk, 1962). One must view an object from the proper aesthetic attitude, which includes displacement of the practical and loss of awareness of self. In assuming this attitude one will be able to show objectively how certain features of the work cause certain emotions (Beyer, 1974). Contrasted to this view is the position that the meaning an object has is related to one's interest in viewing the object (Beyer, 1977).