HUMAN FACTORS, 1997, 39(2), 323–328

COMMENTARY

Heeding the Legacy of Meister, Brunswik, & Gibson: Toward a Broader View of Human Factors Research

KIM J. VICENTE,[1] *University of Toronto, Ontario, Canada*

THE PROBLEM

> The fact that one has to justify a study of human behaviour in a natural setting, where humans ordinarily behave, as opposed to in a laboratory setting, is certainly extraordinary.
>
> Xiao (1994, p. 6)

For years Meister has said that human factors researchers should pay more attention to the characteristics of the operational settings to which their research is intended to generalize; otherwise they would fail to contribute to the solution of applied problems (e.g., Meister, 1989). Decades before Meister, Brunswik (1956) told psychologists that experiments should be representative of the situations, tasks, and stimuli that people encounter in their ecology; otherwise basic research findings would not generalize outside the laboratory. After working as an aviation psychologist in World War II, Gibson (1967/1982) wrote, "Putting my education to a practical test was a new education. I discovered that what I had known before did not work. I learned that when a science does not usefully apply to practical problems there is something seriously wrong with the theory of the science" (p. 18). These scholars were drawing attention to the fact that experimental research in human factors (and psychology) should be both informed by and directed toward naturalistic settings.

To what extent has the human factors community paid attention to the statements made by these widely recognized scholars over the past 50 years? This question is prompted by two passages in a recent issue of this journal. In an investigation of visual display design and monitoring, Payne, Lang, and Blackwell (1995) justified the design of their first experiment as follows: "In Experiment 1 we employed a continuous visual monitoring task that is similar to many real-world tasks, such as those performed by air traffic controllers, pilots, and *nuclear power plant operators*" (p. 510, emphasis added). In an investigation of visual display design in process control, Hansen (1995) pointed out a limitation in the design of his first experiment: "In Experiment 1 . . . the subjects were continuously supervising the changes [in the visual displays]. This is hardly ever the case in control rooms" (p. 550).

What can one conclude when two papers published in the same issue of a scholarly, archival publication so blatantly contradict each other? Given the clear conflict between the two quotations, it seems highly likely that one of them is incorrect, and this is in fact the case (see Vicente, Burns, Mumaw, & Roth, 1996). However, the point of this commentary is not to point out who is right and who is wrong, nor is it to argue that the research based on the incorrect characterization of process control operators' job

[1] Requests for reprints should be sent to Kim J. Vicente, Department of Mechanical & Industrial Engineering, University of Toronto, 5 King's College Rd., Toronto, ON Canada M5S 3G8; benfica@mie.utoronto.ca.

should be dismissed. Instead, my points are as follows: (a) There is a glaring contradiction in the assertions made in these two papers; (b) a closer look at how human factors research has tended to be conducted may explain its origin; and (c) most important, this critical self-examination will reveal some useful implications for how to conduct human factors research in the future.

TWO VIEWS OF HUMAN FACTORS RESEARCH

Types of Research

To achieve this important goal of deriving some useful implications for future human factors research, it is useful to distinguish among four kinds of human factors research. The following categorization scheme is not intended to be exhaustive, rigorous, or unique. However, it captures some meaningful distinctions that are useful for the purposes of this commentary.

Type 1: Highly controlled laboratory experiments. These studies use simplified tasks and try to either hold constant or independently manipulate each factor that may be relevant to the phenomenon under study in order to obtain an unconfounded understanding of the effect of each independent variable. A good example is the work of Gould et al. (1987) on reading from paper versus CRT displays, which won the Jerome H. Ely Award for the outstanding paper published in Volume 29 of *Human Factors.*

Type 2: Less controlled but more complex laboratory experiments. These studies do not explicitly try to control for every factor but, instead, present participants with more complex tasks than in Type 1 research. A particularly messy example is the long-term study of interface design for process control conducted by Christoffersen, Hunter, and Vicente (1997).

Type 3: Evaluations conducted in high-fidelity simulators or in the field. These studies try to maximize representativeness with the hope of increasing generalizability to operational settings, but they give up some control as a result. A good example is the investigation of safety parameter display systems for nuclear power plants conducted by Woods, Wise, and Hanes (1982).

Type 4: Qualitative, descriptive field studies. This type of study is concerned with observing and documenting behavior in the field. Usually no manipulations are conducted, and the results obtained are primarily qualitative in nature. A recent example is the Vicente et al. (1996) study of operator monitoring in nuclear power plants.

The Traditional View

How have these types of research been traditionally viewed in the human factors community? It is always difficult to characterize people's attitudes toward broad issues without being accused of presenting a "straw man" position. To sidestep this criticism, I will deliberately present a caricature of what I perceive to be the traditional view.

Type 1: Highly controlled laboratory experiments. In the traditional view, this is "real" science because it is the only reliable way to discover fundamental principles that are pure in the sense that they are not tied to any particular context. In the words of Banaji and Crowder (1991), "If you wish to do research that is useful (i.e., practical, functional) the *optimal* path is controlled experimentation" (p. 79).

Type 2: Less controlled but more complex laboratory experiments. These are poorly designed experiments. Because some factors are not meticulously controlled for, the results are confounded. This view was held by the reviewer of a paper I wrote recently, who harshly criticized my research for not controlling for *all possible* relevant factors.

Type 3: Evaluations conducted in high-fidelity simulators or in the field. In the traditional view, these studies are not conducted in a "pure" manner and thus will not lead to the discovery of generalizable laws. In the words of Kelso (1995), "most naturalistic behavior is too complicated to yield fundamental principles" (p. 32).

Type 4: Qualitative, descriptive field studies. This research is not scientific because no factors are manipulated by the experimenter (orthogonally or otherwise). And because "hard" quantitative data are not obtained, statistical tests cannot be performed, and thus any findings are considered to be highly subjective and completely speculative.

Critical Analysis

The traditional view is at odds with the insights of Meister, Brunswik, and Gibson because it associates good research solely with highly controlled laboratory experiments. This is a very narrow view of science. Holton (1988), a well-known historian of science, has observed and criticized this view: "The younger sciences . . . are now (erroneously, in my opinion) trying to emulate the older physical sciences by restricting their area of investigation, even if artificially, to . . . phenomenic (empirical) and analytical statements" (p. 3). But as Holton's rich case studies illustrate, science is much more than this, encompassing naturalistic observation, qualitative description and categorization, inductive leaps of faith, and axioms that can never be empirically tested. These facets of science are not respected—and thus not widely practiced—in the traditional view. As a result, this view does not accommodate some of the most important achievements in the history of science, including Einstein's special and general theories of relativity, Darwin's theory of evolution through natural and sexual selection, and Newton's first law of motion. None of these seminal contributions to science resulted directly from controlled experimentation; in fact, the third example cannot be experimentally tested because there is no such thing as an object without any force acting on it (Poincaré, 1905/1952)!

This analysis may explain the origin of the two conflicting quotations cited earlier. Because human factors researchers focus primarily on Type 1 research, they rarely go into the field. As a result, they can only speculate as to the nature of the demands people actually face. In one of the papers cited earlier, this speculation was far off the mark. Is there a different view of human factors research that could avoid these and other problems plaguing the discipline?

An Alternative View

An alternative view of human factors research has existed for years (e.g., Sheridan & Hennessy, 1984). This is a view of science that is broader than the traditional view and is held by prominent scientists in other fields. Lorenz (1973), a Nobel laureate, observed that naturalistic observation is a legitimate scientific activity that should precede formalization, quantification, and controlled experimentation. Before one can meaningfully formalize, quantify, or experiment, one should identify a natural phenomenon that is worth investigating in more detail and categorize the dimensions of that phenomenon to learn what factors should be manipulated experimentally. Every science goes through this process from qualitative observation to quantitative formalization. Human factors researchers tend to forget this, given that older sciences, such as physics, went through this phase hundreds of years ago. Lorenz's insight—that different types of research serve alternative purposes—can be used to create an overarching model that links complementary forms of research into one coherent scheme. The remainder of this section describes this alternative view of human factors research, beginning with different interpretations of the four types of research described earlier.

Type 1: Highly controlled laboratory experiments. From the alternative view, this type of research is useful for comparing competing theories to determine which provides the most accurate account of the topic being studied. To maintain experimental control, however, it is necessary to simplify the situation by omitting factors that are present in operational settings, holding constant variables that vary in operational settings, and making orthogonal variables that are correlated in practice. This leads to a loss of representativeness (Cook & Campbell, 1979). As a result, the findings are unlikely to generalize readily to operational settings (Chapanis, 1967).

The alternative view also differs from the traditional view in terms of how "pure" Type 1 research should be. It is well known that human behavior is very sensitive to task and situational factors (Payne, Bettman, & Johnson, 1993). Thus the alternative view sees context as an important factor to include in the design of controlled experiments, not something that must be stripped away to discover universal laws of behavior. Otherwise "the artificial situation created for an experiment may differ from the everyday world in crucial ways. When this is so, the results may be irrelevant to the phenomena that one would really like to explain" (Neisser, 1976, p. 33). It is ironic that this would lead to findings that are *specific* to the experimental setting, not *universal.*

Type 2: Less controlled but more complex laboratory experiments. From the alternative view, Type 2 research attempts to sacrifice some experimental control for an increase in representativeness. Because the experiment is more representative of operational settings, the likelihood of generalization increases. This statement is a statistical fact, not a matter of opinion. Generalization from a sample to a population is warranted only when the sample is randomly selected from the population of interest. This is why so much attention is paid to randomly sampling participants. People are different, so it is important to include a number of individuals and to select them in an unbiased manner. But as Brunswik (1956) pointed out, the same logic should be, but rarely has been, applied to task and situation variables, despite the fact that they can account for a large proportion of the variance in behavior.

Type 3: Evaluations conducted in high-fidelity simulators or in the field. This type of research sacrifices even more experimental control in order to determine whether results obtained under research of Types 1 and 2 prevail in the face of myriad additional factors that had not been addressed or were held constant. Contrary to the traditional view, field research has shown that Type 3 findings can be meaningfully generalized across a diverse set of application domains when context is explicitly taken into account (Xiao, 1994). Thus this type of research is viewed as legitimate research, not as messy, applied "racehorse" tests with ungeneralizable results.

Type 4: Qualitative, descriptive field studies. From the alternative view, field studies provide researchers with an understanding of the pressing, significant problems that are in need of research. Type 4 research can also provide an idea of which factors should be manipulated in laboratory experiments and which measures should be adopted to evaluate performance.

A Model for Human Factors Research

Figure 1 shows how the different types of research just described can be linked. The thick line is a continuum of research types, with the left pole indicating strictly controlled but unrepresentative studies (Type 1) and the right pole indicating representative but uncontrolled field studies (Type 4). Research Types 2 and 3 fall between these two poles because they move in the direction of decreasing experimental control and increasing representativeness.

The alternative view suggests that the research types can be linked in the order 4, 1, 2, 3. Field studies can be used to observe behavior in situ to identify phenomena that are worth studying under more controlled conditions. Laboratory studies can then be conducted under more controlled conditions in order to develop causal explanations for the observed phenomena. The generalizability of these causal explanations can then be tested under more representative conditions by conducting experiments that are more complex in nature. Finally, a theory or design intervention can be evaluated in high-fidelity simulators or in the field in the presence of a wide range of factors that had been controlled for or eliminated in the laboratory to see whether the same results are obtained.

Several caveats are in order. First, this is not the only way in which human factors research can—or should—be conducted. The model illustrated in Figure 1 is only one coherent way in which to address the challenges posed by human factors research. The model benefits from a basis

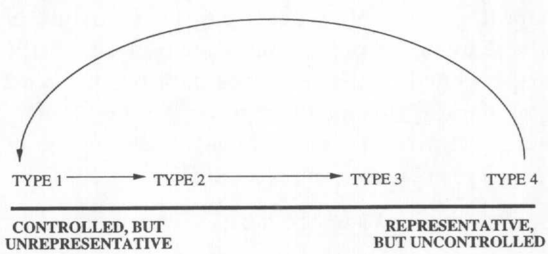

Figure 1. An alternative model for defensible, generalizable human factors research.

in a broad view of science and from exploiting the insights of a diverse set of noted scholars. Second, the research process need not follow the idealized, linear flow shown in Figure 1. Frequently iteration may be necessary (e.g., if Type 3 studies indicate a weakness with a theory, then it may be useful to conduct more Type 1 studies). Third, it is not essential that all these steps be performed by the same researcher.

What *is* important is that all the steps should be addressed in a particular topic of research. If this is not done, then some of the insights that are required for effective technology transfer may not be incorporated. For example, if Type 4 research is not conducted, how else would one meaningfully decide which topics are worth researching and which factors should be investigated experimentally? If one does not conduct Type 1 research, then it is difficult to build theories that have solid empirical support; and without a defensible theory, generalization from one application domain to another would be tenuous.

If Type 2 research is bypassed, then one may be missing a valuable opportunity to evaluate the generalizability of a set of findings under more representative conditions in a relatively economical fashion. Going straight from Type 1 research to costly and time-consuming Type 3 research might be an expensive mistake if a problem is uncovered that could have been identified with Type 2 research. Finally, if Type 3 research is overlooked, then one may introduce an idea or product into an operational setting without being cognizant of the impact that it will have under those conditions. In safety-critical systems, this could lead to disastrous consequences.

In summary, the approach in Figure 1 achieves generalizability and defensibility in the same research program, thereby heeding the lessons of Meister, Brunswik, and Gibson. (For an example of the application of this model from beginning to end in the same research program, see Harrison, 1996.)

CONCLUSION

The discipline of human factors has been in a crisis for some time, with a gap between the basic and applied worlds (Chapanis, 1967; Meister & Farr, 1967; Rouse, 1985). The two blatantly conflicting statements quoted earlier are a symptom of this crisis. Although human factors researchers have devoted much attention to controlled laboratory experiments, these experiments have rarely been directly motivated by field observations. Also, the generalizability of results obtained in the laboratory has rarely been evaluated under more representative settings. The implications of these choices were noted by Meister, Brunswik, and Gibson. I believe that these warnings have gone relatively unheeded because the field of human factors has had a narrow view of research. Furthermore, if one's view of science cannot account for the achievements of Newton, Darwin, and Einstein, then there is a problem. If human factors researchers broaden their view, as suggested earlier, then the prospects for making a difference in the applied world will be much greater.

It is important to reemphasize that these ideas are not new. People far more knowledgeable and experienced than I have been saying essentially the same thing for many years. Judging from the research published in this and other journals in the field, however, very few have been listening. The message is worth repeating because the perils of ignoring it are too great.

ACKNOWLEDGMENTS

The writing of this article was sponsored by grants from the Natural Sciences and Engineering Research Council of Canada. I thank Bill Howell for giving me the opportunity to write this commentary and for his helpful suggestions, and an anonymous reviewer for providing balanced and constructive comments.

REFERENCES

Banaji, M. R., & Crowder, R. G. (1991). Some everyday thoughts on ecologically valid methods. *American Psychologist, 46,* 78–79.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Chapanis, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics, 10,* 557–577.

Christoffersen, K., Hunter, C. N., & Vicente, K. J. (1997). A longitudinal study of the effects of ecological interface design on fault management performance. *International Journal of Cognitive Ergonomics, 1,* 1–24.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton-Mifflin.

Gibson, J. J. (1967/1982). James J. Gibson: Autobiography. In E. Reed & R. Jones (Eds.), *Reasons for realism: Selected essays of James J. Gibson* (pp. 7–22). Hillsdale, NJ: Erlbaum. (Reprinted from *History of Psychology in Autobiography*, Vol. 5, Ed. E. G. Boring, 1967, Irvington)

Gould, J. D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., & Minuto, A. (1987). Reading is slower from CRT displays than from paper: Attempts to isolate a single-variable explanation. *Human Factors, 29,* 269–299.

Hansen, J. P. (1995). An experimental investigation of configural, digital, and temporal information on process displays. *Human Factors, 37,* 539–552.

Harrison, B. L. (1996). *The design and evaluation of transparent user interfaces: From theory to practice.* Unpublished doctoral dissertation, University of Toronto, Ontario, Canada.

Holton, G. (1988). *Thematic origins of scientific thought: Kepler to Einstein* (Rev. ed.). Cambridge: Harvard University Press.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge: MIT Press.

Lorenz, K. Z. (1973). The fashionable fallacy of dispensing with description. *Die Naturwissenschaften, 60,* 1–9.

Meister, D. (1989). *Conceptual aspects of human factors.* Baltimore: Johns Hopkins University Press.

Meister, D., & Farr, D. E. (1967). The utilization of human factors information by designers. *Human Factors, 9,* 71–87.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* New York: Freeman.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge, England: Cambridge University Press.

Payne, D. G., Lang, V. A., & Blackwell, J. M. (1995). Mixed versus pure display format in integration and nonintegration visual display monitoring tasks. *Human Factors, 37,* 507–527.

Poincaré, H. (1905/1952). *Science and hypothesis.* New York: Dover. (Reprinted from *Science and hypothesis,* 1905, Walter Scott Publishing Company)

Rouse, W. B. (1985). On better mousetraps and basic research: Getting the applied world to the laboratory door. *IEEE Transactions on Systems, Man and Cybernetics, SMC-15,* 2–8.

Sheridan, T. B., & Hennessy, R. T. (1984). *Research and modeling of supervisory control behavior.* Washington, DC: National Academy Press.

Vicente, K. J., Burns, C. M., Mumaw, R. J., & Roth, E. M. (1996). How do operators monitor a nuclear power plant? A field study. In *Proceedings of the 1996 American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies* (pp. 1127–1134). LaGrange Park, IL: American Nuclear Society.

Woods, D. D., Wise, J. A., & Hanes, L. F. (1982). *Evaluation of safety parameter display concepts* (Vol. 1; Final Rep. NP-2239). Palo Alto, CA: Electric Power Research Institute.

Xiao, Y. (1994). *Interacting with complex work environments: A field study and a planning model.* Unpublished doctoral dissertation, University of Toronto, Ontario, Canada.

Kim J. Vicente received a Ph.D. (1991) in mechanical engineering from the University of Illinois at Urbana-Champaign. He is an associate professor of mechanical and industrial engineering and of biomedical engineering and director of the Cognitive Engineering Laboratory at the University of Toronto.

Date received: April 23, 1996
Date accepted: November 25, 1996