N88-23383

5/3-53
140801
18P.

Defining and Measuring Pilot Mental Workload

Barry H. Kantowitz
Battelle Memorial Institute
Human Affairs Research Centers
Seattle, Washington 98105

BE 282249
CN

Both scientists and practitioners agree that definition is a necessary precursor to productive discourse. But any definition must be clearly understood by both parties. For example, the hip musician's definition of jazz --Jazz is when you dig it, man!--does not help the naive listener who sincerely wants to appreciate jazz music but lacks the artistic sophistication of the professional musician. While this definition of jazz is too simple, the musician can also confuse a listener by excessive use of jargon that is too sophisticated. Few listeners could sympathize with a jazz trumpet player who complained about being boxed in by a C minor ninth vamp laid down by his pianist.

Similar dangers abound when research scientists try to define and explain mental workload to airplane pilots and other interested non-researchers. As a researcher I am well aware that the jargon used by human factors specialists may not always make sense to the uninitiated. Yet I also understand that an overly simple definition of mental workload --Too much mental workload is when you can't fly the plane right --also is not helpful. My goal in this article is to try to explain to the pilot why and how workload researchers approach what may appear to the pilot as a simple problem in very complex ways. There just is no easy way to define and measure mental workload.

Why Use Theory?

Researchers and practitioners can be arranged along a hypothetical continuum according to how they approach solving a problem. At the cost of only minor exaggeration we might characterize practitioners as being so anxious to solve a problem that they often solve the wrong problem whereas researchers are so anxious to get everything right that they seldom solve any problems! In order to reach a satisfactory solution, albeit not necessarily an optimal one, we must operate nearer to the middle of this continuum instead of at an extreme endpoint. It is true that an experienced problem-solver can often come up with a satisfactory answer without explicitly invoking theory. But I would argue that this approach is too idiosyncratic to work in general. The world does not have enough experienced problem solvers to meet every need. However, one theory goes a long way. It can be applied to many different practical scenarios. Theories offer generality. We do not need a separate theory for each problem. We may not even need a very complex theory to get a direction for solving a practical problem like evaluating pilot mental workload. After all, you don't need a Ferrari to go grocery shopping. A Volkswagen will get you to the store and back. When I am asked to solve a problem like measuring pilot mental workload, I start out by looking for a handy theory. I do not expect the theory to solve my problem, only to get me started in a promising direction. Theory can be a filter that narrows down a large set of possible approaches allowing us to concentrate our efforts upon a few techniques that are most likely to yield satisfactory solutions.

There is a deplorable tendency for the practitioner to avoid theory because it does not seem relevant to the immediate problem at hand. Each problem is seen as an isolated issue and, practitioners who avoid theory run the considerable risk of reinventing the wheel time and time again without realizing it. But even the practitioner who wants to use theory must face at least two major obstacles. Most psychological theories have been formulated in arcane ways with little regard for fostering practical applications. Furthermore, there are too many theories so that it is hard for the practitioner to select one theory from the abundance created by diligent researchers. Later on I will suggest one particular kind of theory that should be useful for studying pilot mental workload. For now, I acknowledge these obstructions.

I believe that theory offers four substantial benefits to the practitioner faced with a real-world problem. First, it fills in where data are lacking. We will never have enough empirical results to solve all problems. Theory is needed for accurate and sensible interpolation. Second, theory can yield the precise predictions that engineers and designers demand. It is better to have predictions about the workload imposed on a pilot by some particular system design than to have to build the system and then obtain data to fix the next version. Third, theory prevents us from reinventing the wheel. It allows us to recognize similarities among problems. Fourth, theory is the best practical tool. Once an appropriate theory is available, it can be used cheaply and efficiently to aid system design.

Limited Capacity Theory of Attention

My approach to the practical problem of pilot mental workload is derived from basic research on attention. A detailed analysis of the kind of theory best suited for this work can be found in Kantowitz (ref.1). Here I will only summarize my conclusions in this regard. I prefer an attention theory with a single limited pool of capacity as the starting point for studies of pilot mental workload. Such a model was popularized by Broadbent (ref.2). While current views of attention realize that many of the details of this original limited-channel model are incorrect (see ref. 3 for a review), the fundamental idea of a single limited-capacity source that funds mental operations remains sound. This concept of attention is particularly useful for work on pilot mental workload because it carries with it the idea of *spare capacity*. Spare capacity is roughly defined as extra capacity not currently being used by the human but available immediately should the need arise.

There are certain assumptions used by most basic researchers studying attention and capacity that deserve explicit mention (ref. 3). First, we assume that behavior can be understood in terms of a hypothetical flow of information inside the organism. This flow cannot be directly observed but must instead be inferred from overt measures of performance. Models must not only duplicate the overt performance but must also make reasonable statements about this postulated internal information flow. For example, a female singer and a tape recording made with the proper brand of tape can both shatter a slender crystal goblet. Nevertheless, no one would claim that the human vocal tract and an electronic tape recorder produce sound by the same internal information flow.

Second, we assume that capacity is the "price" each internal processing stage charges the system to perform its own activity or information transformation. If sufficient capacity is not available, the internal processing stage may be unable to perform its function properly and/or may require greater processing time.

Third, we assume that allocation rules determine how capacity is mapped to internal stages. This is especially important when demand exceeds supply. A complete model of attention and information processing should have something explicit to say about each of these three key assumptions (ref. 3).

Defining Mental Workload

Mental workload is an intervening variable, similar to attention, that modulates or indexes the tuning between the demands of the environment and the capacity of the organism. Before considering the implications of this definition I must first explain what I mean by "intervening variable."

Intervening variables have been the subject of much discussion in psychology, especially as contrasted with hypothetical constructs (ref. 4). A hypothetical construct has surplus meaning; for example, one might try to locate the physiological basis of the hypothetical construct called the limited-capacity channel. An intervening variable is closely coupled to the operations that define it. Indeed, it ceases to exist without these operations. For example, learning is often defined as a relatively permanent change in behavior between the first test of some knowledge and a later test. Presumably better performance on the later test is evidence for the intervening variable we call learning. If the tests are removed, we can no longer make any statements about learning. Learning is thus inferred from a change in performance. It cannot be observed directly.

In a similar manner, both attention and mental workload are also intervening variables. They cannot be observed directly. We make inferences about attention or workload only on the basis of observed changes in performance. If performance decreases we often attribute this decrease to increased mental workload (or decreased attention).

There are at least four important implications of the definition of mental workload stated above. First, it implies that both underload and overload are cause for concern. In both cases there is an imbalance between the demands of the environment and the capabilities of the organism. A crew falling asleep on a trans-oceanic flight is as much a pilot mental workload problem as an engine fire. Second, the definition implies that capacity is fixed. Third, to be most useful the definition implies that spare capacity is related to mental workload and this in turn implies that a single-pool model of capacity will work better than attention models that postulate multiple sources of capacity. Fourth, it implies that the limit upon the internal information flow within the human is one of rate not amount. An analogy (ref. 5) will make this clear. No highway engineer is truly interested in the number of cars that a freeway can hold as a static measure. While this number is important for designing parking lots, highway engineers are far more concerned with the number of cars that can flow past a given point in some specified time. Similarly, the amount of information per unit time, bits/sec, that can flow through the human is more important for understanding pilot mental workload than an absolute amount of information with no time constraint.

Measuring Mental Workload

There are three general methods for measuring pilot mental workload: (1) subjective measures, (2) objective measures, especially those based upon secondary tasks, and (3) psychophysiological measures. These are discussed in general by Kantowitz (ref. 1) and as they relate to aviation by Kantowitz and Casper (ref. 6). All methods have advantages and disadvantages. There is no clearly superior method to measure pilot mental workload in all circumstances. I believe that secondary-task measures offer the best opportunity to obtain valid and reliable indices of pilot mental workload now. In the near future psychophysiological measures may also prove to be quite useful.

The reader may be surprised that I have not endorsed subjective measures, since these are by far the most widely used method at present. While it is awfully easy to obtain subjective measures, they are quite difficult to interpret. There are at least two fundamental problems with them. First, with the possible exception of SWAT* ratings (ref. 7), the psychometric properties of most subjective rating scales have not been established. While at least interval scale properties are required for meaningful measurement and comparison, it is not at all clear that more than ordinal measurement has been achieved in most cases. Second, people are not very good at giving direct introspections that accurately reflect their own internal mental states. Psychology has long abandoned the method of introspection because it utterly failed to produce reliable data. A more recent example can be found in the work of Metcalfe (ref. 8) who studied people's ability to solve anagram puzzles and other brain teasers. Every ten seconds subjects were asked to rate on a scale of 0 to 10 how close they felt they were to a correct solution. The results were extremely lucid. People were grossly inaccurate in their ratings. When they gave high ratings, indicating that they thought they were close to a correct solution, they were more likely to give an incorrect answer than to reveal the proper solution. This demonstrates once again that subjective intuitions may not

*Subjective workload assessment technique (SWAT) be reliable.

Thus, we are better off relying upon objective data provided by secondary tasks and psychophysiology. The secondary-task paradigm attempts to obtain direct estimates of spare capacity, and hence mental workload, by requiring an additional task to be performed at the same time as the primary flying task. Decrements in secondary-task performance are interpreted as reflecting mental workload imposed by the primary task. Primary tasks that demand greater mental workload will cause poorer performance on the concurrent secondary task.

In order for this interpretation to be valid, several control conditions must be included in the experimental evaluation of mental workload; see Kantowitz (ref. 3) for a detailed explanation and examples of published research where these safeguards have been neglected. The crucial assumption of the secondary-task method is that insertion of the secondary task does not alter primary-task performance or the internal information flow within the human operator.

In the past, secondary tasks were chosen largely on the basis of convenience with little thought given to the theoretical or methodological implications of secondary-task selection. Now, however, it is generally realized that there is no panacea that will create a universal secondary task. Many issues must be considered carefully before a satisfactory secondary task can be accomplished. Some relevant questions are:

1. Will this research be carried out in [1] an operational setting [2] a flight simulator [3] a laboratory?

2. The primary task is [1] flying [2] tracking [3] other continuous task [4] other discrete task.

3. Most primary-task information is presented [1] visually [2] auditorally [3] tactually.

4. The primary-task input information load (e.g., rate of information per unit time such as bits/sec) is [1] low [2] medium [3] high.

5. Input information load is [1] constant [2] low variability [3] high variability.

6. Output modality is mostly [1] manual [2] verbal.

7. Output responses occur [1] seldom [2] moderately often [3] frequently.

8. Operators are [1] unpracticed [2] moderately practiced [3] highly practiced professionals.

9. Operator motivation is [1] low [2] moderate [3] high.

10. Procedures associated with the primary task are [1] well-specified and usually performed in a consistent manner [2] leave the operator some discretion for arranging his work [3] vague and subject to considerable interpretation.

These considerations are sufficiently complex so that an expert system is now under construction to help choose appropriate secondary tasks. Workload COnsultant for Secondary Task Selection (W. COSTS) presents lists of questions similar to those above and makes recommendations for selecting suitable secondary tasks. This expert system uses rule-based chaining to derive its suggested secondary tasks (ref. 9).

A Simulator Example of Secondary-Task Research

At the risk of appearing immodest I will illustrate secondary-task techniques with a series of studies my colleagues and I have conducted in a motion-base (GAT) flight simulator at Ames Research Center (refs. 10,11,12 and 13). The primary task in all these studies was flying the simulator. The secondary task was choice-reaction time with two, three, or four alternatives. This contrasts with the typical study where a simple (one-choice) secondary reaction task has been used. However, based upon a hybrid model of attention (ref. 14) I believed that simple probe tasks were too insensitive and subject to a host of methodological problems. While many researchers felt it would be safer to use a simple probe task because this simple task would be less likely to interfere with the primary flying task, I disagreed. I believed that professional pilots would not allow the secondary task to interfere with flying. The first responsibility of a pilot is to keep the airplane safely in flight. Therefore, professional pilots seemed to me to be the ideal population for taking the risks associated with a complex choice-reaction secondary task.

Results have been excellent. Flying performance measured by root mean square error was not adversely affected by adding the complex secondary task. Furthermore, this secondary task was able to discriminate among levels of workload in many different simulated flight situations. I conclude that the choice-reaction task should be high on everyone's list of preferred secondary tasks. Indeed, this opinion of mine is reflected in W. COSTS which tends to suggest choice reactions for almost any situation where pilot mental workload must be measured.

Psychophysiological Measures

Objective measures need not be only behavioral. The technology for recording psychophysiological correlates of behavior is now well advanced and many of these biological indicants have been used to estimate pilot mental workload (ref. 15). Once monitoring electrodes have been attached to the pilot, these indices have the advantage of being relatively unobtrusive. They do not interfere with flying as might be the case for behavioral secondary tasks. However, these data are often difficult to interpret even though they are easier to understand than most subjective ratings. Theories of psychophysiology are not yet as advanced as theories of attention and do not provide a complete framework for interpreting data.

In my laboratory we have had modest success in using heart rate (sinus arrhythmia) and evoked potential as indicants of attention in a psychological refractory period task (ref. 16) and a divided attention task described later in this volume (ref. 17). Others have successfully used psychophysiological tasks to measure pilot mental workload (see ref. 6 for a review). I believe that as theoretical models of psychophysiological indicants are refined, these techniques will become an important part of the toolbox used by human factors specialists to measure pilot mental workload.

Conclusions

The best practical tool is a good theory. Models of attention based upon a single pool of limited capacity offer an excellent starting point for measuring pilot mental workload. Thus, I define mental workload as an intervening variable similar to attention.

Objective measures are preferable for measuring pilot mental workload. Secondary tasks, especially choice-reaction time, are extremely useful in this regard. Psychophysiological tasks will be more useful in the near future as theoretical models are refined.

Acknowledgment

This research was supported by Cooperative Agreement NCC 2-228 from the National Aeronautics and Space Administration to BITS, Inc. S.G. Hart was the NASA Technical Monitor.

References

1. Kantowitz , B.H. Mental workload. In P. Hancock (Ed.), Human Factors Psychology. Amsterdam: Elsvier, 1987.

2. Broadbent, D.E. Perception and communication. Oxford: Pergamon Press, 1958.

3. Kantowitz , B.H. Stages and channels in human information processing: A limited analysis of theory and methodology. Journal of Mathematical Psychology, 29, 1985, 135-174.

4. MacCorquodale, K. & Meehl, P. On a distinction between hypothetical constructs and intervening variables. Psychological Review, 55, 1948, 95-107.

5. Broadbent, D.E. Application of information theory and decision theory to human perception and reaction. In N. Wiener & J.P. Schade (Eds.), Cybernetics of the Nervous System. Amsterdam: Elsvier, 1965.

6. Kantowitz , B.H. & Casper, P.A. Human workload in aviation. In E. Wiener & D. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press, 1987.

7. Reid, G.B., Eggemeier, T., & Shingledecker, C.A. Subjective workload assessment techniques. In M.L. Frazier & R.B. Crombie (Eds.), Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics. Edwards Air Force Base, Ca: AFFTC-TR-82-5, 1982.

8. Metcalfe, J. Premonitions of insight predict impending error. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1986, 12, 623-634.

9. Kantowitz , B.H. Workload Consultant for Secondary Task Selection. Interim report, BITS, Inc., Dec 1987.

10. Kantowitz, B.H., Hart, S.G., & Bortolussi, M.R. Measuring pilot workload in a moving-base simulator: I. Asynchronous secondary choice-reaction task. Proceedings of the Human Factors Society, 27, 319-322, 1983.

11. Kantowitz , B.H., Hart, S.G., Bortolussi, M.R., Shively, R.J. & Kantowitz , S.C. Measuring pilot workload in a moving-base simulator: II. Building levels of workload. Proceedings of the Annual Conference on Manual Control, 20, 1984, 359-371.

12. Bortolussi, M.R., Kantowitz , B.H., & Hart, S.G. Measuring pilot workload in a motion base trainer. Applied Ergonomics, 17, 1986, 278-283.

13. Kantowitz , B.H., Bortolussi, M.R. & Hart, S.G. Measuring pilot workload in a motion base simulator: III. Comparing synchronous and asynchronous secondary choice-reaction tasks. Proceedings of the 31st Annual Meeting of the Human Factors Society, 1987.

14. Kantowitz, B.H. & Knight, J.L. Testing tapping timesharing: II. Auditory secondary task. Acta Psychologica, 40, 1976, 343-362.

15. Coles, M.G.H., Donchin, E. & Porges, S.W. (Eds.) Psychophysiology. New York: Guilford Press, 1986.

16. Kantowitz, B.H. Biocybernetic analysis of a hybrid workload model. Behavioral Research Associates Technical Report BRA 85-10, May 1985.

17. Kantowitz, B.H. & Casper, P.A. Estimating the cost of mental loading in a bimodal divided-attention task: Combining reaction time, heart-rate variability and signal detection theory. NASA CP 2504, pp. 211-229, May 1988.