

An evaluation of computer-based driving systems for research, assessment, and advisement

DORIS AARONSON

New York University, New York, New York

and

JOHN EBERHARD

National Highway Traffic Safety Administration, Washington, D.C.

Evaluations are made of three computer-based driving systems presented at the 1993 Society for Computers in Psychology conference. These systems are compared and contrasted on the dimensions of realism, validity, cost, data-output, and application goals. All three systems have positive attributes on each of the five dimensions, but in different ways, yielding systems that complement each other in their usefulness for research, assessment, and advisement. We note that the systems have strong potential for applications with elderly drivers, but that all three systems need expansion for use with high-risk youthful drivers.

The three computer-based driving systems demonstrated in the present symposium by Schiff, Arnone, and Cross (1994), McKnight and McKnight (1994), and Gianutsos (1994) all have strong positive attributes. As they are each based on somewhat different philosophies and involve different types of hardware and software, their individual merits vary somewhat on several dimensions. Thus, we evaluate these systems below on five dimensions: realism, validity, cost, data-output, and application goals.

Realism

All three contributors were concerned about the realism of their computer-based driving systems, but in relation to their multifaceted goals. In general, they have asked themselves (a) what the *minimum* amount of realism is that is necessary to accomplish the goals, and (b) *where* in the total system that realism should be located. They have each arrived at different answers, and thus the three systems complement each other in interesting ways.

Schiff et al. (1994) have incorporated realism in four ways. First, they have developed a videotape- (or alternatively videodisc-) based system in which subjects view a realistic environment shot from a camera positioned in a moving automobile near the driver's head. Second, each of their scenarios includes a complex of visual factors demanding a complex of psychological reactions, as Schiff et al. feel (a) that there are numerous individual driving strategies consisting of compensatory components interacting in complex ways, and (b) that real-world driving situations involve multiple stimulus dimensions interact-

ing in complex ways. Third, they have developed ways to vary some visual factors systematically to reflect real-world variations, for example, using Macintosh video-toaster technology to vary brightness. Fourth, they have two realistic foot pedals for accelerating and braking responses, providing a limited degree of subject interaction with the video displays. They are now in the process of developing software to incorporate a noninteractive steering wheel for response collection.

McKnight and McKnight (1994) have also used the videotape and videodisc technology to accomplish realism. In addition to the forward view, their subjects see a second visual display at the top of the monitor, namely, the display one would see in a rear-view mirror. That can give subjects early warnings of another vehicle that might try to pass. They have purposely chosen to simplify their video displays, so that each scenario primarily emphasizes one psychological factor. They are taking a psychometric approach to driving in order to isolate and study one human factor at a time. Thus, they also have a sizeable number of abstract, laboratory-type displays of squares, digits, moving arrows, etc., to provide traditional indicators of, for example, reaction time, selective attention, and short-term memory. They hope to show correlations between performance on these established indicators of cognitive processing and that on their more realistic videotape scenarios. McKnight and McKnight have chosen a decidedly nonrealistic response device for the subject—a joystick, connected into the computer game port, and which can be deflected in various directions by the subject and has a button on top that can be used for discrete responses. This reduces costs and simplifies programming, compared with more realistic response devices.

Gianutsos (1994) has purposely designed graphic displays based on computer symbols. Although these displays lack video realism, they permit interaction not possible in the

Reprints can be obtained from either D. Aaronson, NYU Psychology Department, 6 Washington Place (8th floor), New York, NY 10003 (e-mail: doris@xp.psych.nyu.edu), or from J. Eberhard, National Highway Traffic Safety Administration, 400 7th St., S.W., Washington, DC 20590. Correspondence should be addressed to D. Aaronson.

other two systems presented. The subject can easily move the symbolic car in various directions down the continuously winding "road." Further, Gianutsos has developed a relatively small number of critical driving tasks in comparison with the much larger sets used by the other presenters. Gianutsos' selections are based on her prior research with neurological patients who have particular cognitive and perceptual impairments that affect safe driving. This task selectivity makes possible a quick and efficient assessment system for *individuals* who may be at high risk on the road. In contrast, the other two presenters have developed large "batteries" of items that should show correlations with real-world driving for a *general population*. Whereas the other presenters have developed realism on the visual side, Gianutsos has built in considerably more realism on the response and perceptual-motor interaction front. Her subjects can control the direction of the graphic car with a realistic steering wheel and foot pedal, and can signal potential right or left turns with a realistic blinker toggle switch while driving.

Thus, we have an important set of relatively unanswered questions as to *where* in the system realism is important, and *why*, in terms of (1) psychological mechanisms and (2) predictive validity. It will be interesting to see the data from each of the three approaches over the next few years.

Validity

The three systems also take slightly different approaches toward demonstrating validity. Schiff et al. (1994) and McKnight and McKnight (1994) have high "face validity" in their visual displays. They are also aiming for predictive validity by correlating performance on a composite set of scenarios with indices of driver performance (e.g., crash records). But such validity will be difficult to establish because of the relatively low incidence of crashes in the population. However, Schiff et al. have made empirical progress in this direction with a pool of over 170 subjects, many of whom are elderly drivers. They have obtained moderate correlations ($.40 < r < .47$) between subjects' performance and a composite "risk scale" including crash information.

McKnight and McKnight (1994) also aim to base some of their system's validity on the inclusion of numerous established laboratory-type tasks which past psychological articles have shown to be correlated with driving performance. Since both Schiff et al. (1994) and McKnight and McKnight have many scenarios and/or tasks, some of them will likely prove to have poor predictive validity as individual components and may ultimately be removed from the assessment battery. Large subject groups will be needed for them to end up with a smaller battery that (1) can be standardized and administered in a reasonable amount of time, (2) includes a *set* of items with combined predictive validity, and (3) also includes items that can implicate particular environmental and psychological mechanisms involved in good or poor driving performance.

Gianutsos (1994) has provided fairly strong empirical evidence for the validity of her system by showing per-

formance data that discriminate among populations known to vary in driving ability on the basis of *independent* indicators (i.e., concurrent validity). Specifically, four different simulation performance scores show statistically reliable variations among four populations: (1) a normalization group of working-age drivers; (2) neurological patients who failed other types of driver evaluations, including road tests; (3) neurological patients who passed such alternate evaluations; and (4) a large pool of AARP elderly drivers from a Hartford Insurance Company study.

Costs

All three systems, including their computers, cost in the region of \$5,000. Thus, all developers should be commended for developing useful systems at a cost that is quite realistic in terms of widespread adaptation in research, assessment, or advisement contexts (e.g., motor vehicle departments and hospital clinics). In contrast, other currently available systems cost around \$50,000.

Further, all three systems appear to be relatively "user-friendly." Thus, new people will be able to use them in various types of applications with only a day or so of training, together with access to a telephone hot line for occasional questions or problems. Thus, cost measured in time (as opposed to dollars) is also economical for all three systems. Finally, one can measure cost in terms of safety. A simulation system is safer for both driver and tester than a real-world road test—particularly for populations that might be at risk for driver safety, such as neurological patients, some elderly people, and the visually impaired.

Data Output

Again, all three system developers can be commended for providing rather complete and comprehensible outputs of response times and errors for each simulated driver. These are in the usual form of captioned statistical tables and spreadsheets that professional researchers are used to reading. Thus, all three systems will provide data in formats with which the scientific community is comfortable. This will enhance the application of these systems as research tools.

However, Gianutsos (1994) has gone well beyond that in also providing both computer displays and hard-copy output that can be interpreted easily by the typical driver or tester who does *not* have a technical background. Three characteristics of these outputs are important. (1) Performance scores are compared with subjects' prior self-appraisal on each driving component, yielding important "metacognitive" indices (i.e., the extent to which subjects under- or overestimate their own driving abilities). (2) Performance on these particular outputs is not given in absolute reaction time or error units, but rather is based on relative or standardized scores similar to IQ scales, with 100 indicating the mean of the normalization group and standard deviation units spaced every 15 points along the scale. Thus, the subject and the tester both have a familiar frame of reference in which to interpret the individual subject's scores. (3) Finally, the output contains

ample verbal interpretations. For example, all numbers have labels such as "fair" or "average," and English sentences further explain the output (e.g., "an average rating ranges from 85 to 115; below 70 is *unacceptable*; very few safe drivers score that low"). Such an output format can markedly enhance the value of the driving simulator for assessment and advisement.

Goals

All three systems appear to have strong potential for the three goals indicated by the title of this symposium, namely, research, assessment, and advisement. However, each system may have a slight edge over the others on one or another of the goals. McKnight and McKnight's (1994) system may have an edge for some types of research. Their "abstract tasks" may provide a nice bridge between traditional laboratory perceptual and cognitive tasks, on the one hand, and their videotaped driving-related scenarios, on the other. Further, they focus each task on a separate psychological process, and the individual tasks are generally designed with a psychometric or psychophysics approach. Schiff et al. (1994) may have an edge on driver assessment in a broad or general population because of the moderate amount of realism on both the display and the response sides, and because of the variety of scenarios related to known high-crash environments (e.g., high levels of approaching headlight glare, and cross-traffic left turns). Gianutsos (1994) may have the edge on advisement because of her creative output of information for nontechnically trained people and because poor performance on the various procedural phases points directly to psychological mechanisms (e.g., impulsivity, perceptual-motor coordination) that might be changed by cognitive rehabilitation.

We note that one potentially important goal is missing from all three systems—that of *training*. When drivers have shown poor performance on one or another system, what should be done? Indeed, Evans (1991) points out that there is little past evidence to support various types of driver training or driver education. That is to say, comparable groups with and without particular types of training (e.g., defensive-driving courses) show little difference in future crash rates. However, according to Evans, although past training methods have not been successful, other methods might be helpful. He cites Michon's (1989) claims that rule-based approaches grounded in cognitive psychology have promise. We would suggest combining those with motivational approaches grounded in behavioral psychology. In other words, safe cognitively based rules must be turned into reliably executed driving habits.

As a final comment, we note that all three systems were developed especially for use with elderly drivers. Dur-

ing recent years there has been substantial discussion concerning the safety of older drivers, particularly those in their eighties, because of their decreasing vision and cognition. However, a careful look at a wide variety of statistics (see, e.g., Evans, Ch. 2) provides strong evidence that severe crash involvement for drivers aged 60–70 or over differs little from that for drivers aged 40–50. Older drivers often are aware of their limitations, drive less, and self-limit their driving to daylight trips on familiar routes.

In contrast, the highest safety problems are among drivers aged 16–26. Their high crash rates (three to five times that of 40-year-old drivers) cannot be accounted for simply by lack of driving "skills." Rather, there is evidence that much of their problem relates to a high degree of risk-taking behavior, and that young drivers too often experience risk as intrinsically rewarding (Fuller, 1988; Jonah, 1986). Thus, the developers might consider ways to expand their simulation systems to include tasks or scenarios appropriate for *measuring risky driving behaviors* in drivers of all ages. These might include making *choices* as to when to pass another car, how close to follow behind another vehicle, and how fast to travel in heavy traffic or bad weather conditions. These judgmental tasks have goals quite different from Schiff et al.'s (1994) low visibility tasks, McKnight and McKnight's (1994) high information load tasks, and Gianutsos' (1994) choice reaction time tasks.

In sum, we have seen presentations of creative, cost-effective, and useful driving simulation systems. We look forward to seeing new data from the contributors in future years, and to seeing how their systems continue to develop and improve over time.

REFERENCES

- EVANS, L. (1991). *Traffic safety and the driver*. New York: Van Nostrand Reinhold.
- FULLER, R. (1988). Psychological aspects of learning to drive. In J. A. Rothergatter & R. A. de Bruin (Eds.), *Road users and traffic safety* (pp. 527-537). Assen/Maastricht, The Netherlands: Van Gorcum.
- GIANUTSOS, R. (1994). Driving advisement with the elemental driving simulator (EDS): When less suffices. *Behavior Research Methods, Instruments, & Computers*, **26**, 183-186.
- JONAH, B. A. (1986). Accident risk and risk-taking behavior among young drivers. *Accident Analysis & Prevention*, **22**, 97-107.
- MCKNIGHT, A. S., & MCKNIGHT, A. J. (1994). The Automated Psychophysical Test (APT) for assessing age-diminished capabilities. *Behavior Research Methods, Instruments, & Computers*, **26**, 187-191.
- MICHON, J. A. (1989). Explanatory pit falls and rule-based driver models. *Accident Analysis & Prevention*, **21**, 341-353.
- SCHIFF, W., ARNONE, W., & CROSS, S. (1994). Driving assessment with computer-video scenarios: More is sometimes better. *Behavior Research Methods, Instruments, & Computers*, **26**, 192-194.