

USING MICROWORLDS TO DESIGN INTELLIGENT INTERFACES THAT MINIMIZE DRIVER DISTRACTION

Barry H. Kantowitz
University of Michigan Transportation Research Institute
Ann Arbor MI USA 48109-2150

ABSTRACT

While recent developments in telematics have produced great interest in driver distraction, this is hardly a new topic. An early UMTRI report (Treat, 1980) defined internal distraction as a diversion of attention from the driving task that is compelled by an activity or event inside the vehicle. Based on data collected in Monroe County Indiana, Treat (1980) concluded that internal distraction was a factor in 9% of in-depth reports and 6% of on-site investigations. In the period of data collection (1972-1975) conversation with a passenger and increasing use of entertainment tape decks were the major sources of distraction. Now a host of modern infotronic devices offers even greater opportunities for internal distraction (Kantowitz, 2000).

Intelligent driver-vehicle interfaces present a wonderful opportunity to successfully manage this increased in-vehicle workload. This smart interface would be adaptive, making dynamic allocation of function decisions in real time. Designing such an intelligent interface presents many problems. In particular, since new infotronic devices are being developed and deployed rapidly, it seems difficult to evaluate all these new designs. This chapter focuses upon using microworlds to swiftly assess effects of in-vehicle infotonics upon driver distraction.

Microworlds vary along several dimensions such as realism, tractability and engagement (Ehret, Gray, & Kirschbaum, 2000). The traditional driving simulator is only one example of a relevant microworld. By considering a wider range of microworlds, we can gain insight into how to best utilize driving simulators. Issues of validity are also illuminated when considered from a microworld perspective. If appropriate intelligent interfaces are designed, telematics should never increase driver distraction.

INTRODUCTION

Driver distraction, although hardly a new topic, has been much in the public mind recently due to increasing popularity of in-vehicle cell phones and telematics. This is a great opportunity to demonstrate that ergonomic solutions are far more meritorious than legislative solutions. Many localities are considering legislation to control the use of cell phones in moving vehicles. Unfortunately, the modal legislation being considered would ban hand-held phones but not hands-free phones. Since the conversation is a much more important determinant of driver distraction than the dialing (Goodman, Tijerina, Bents, & Wierwille, 1999), such legislation, although perhaps increasing safety immediately because of the great number of hand-held

phones currently used by drivers, would not solve the problem and might make it worse in the long run by encouraging the false belief that using hands-free phones is without risk. I recently testified before the Michigan House Transportation Committee which is considering a bill to increase the penalty for drivers who are using a cell phone when an accident occurs. This is a controversial bill and some legislators are reluctant to impose restrictions that would be more stringent, such as banning phones entirely, without the benefit of on-road accident data regarding driver distraction. Since Michigan has only this year added cell phones to the accident reporting form used by state police, it will be several years before sufficient data are accumulated to allow a judgment about the severity of the problem.

Research on driver distraction is hardly new. An earlier UMTRI report (Treat, 1980) defined internal distraction as a diversion of attention from the driving task that is compelled by an activity or event inside the vehicle. This report was based upon data collected from 13,568 police-reported accidents that occurred in Monroe County Indiana from 1972-1977. It distinguished internal distraction, defined above, from inattention, defined as a noncompelled diversion of attention from the driving task. The study used a tri-level approach to accident investigation. Baseline data were obtained for all 13,568 accidents. This was followed by on-site investigation of 2,258 cases. A subset of 420 cases were investigated in depth by a multidisciplinary team. Internal distractions were causal factors in 9% of in-depth reports and 6% of on-site investigations. This compares with inattention as a causal factor in 15% of in-depth reports and 14% of on-site investigations. In those days there were no cell phones inside vehicles and the two main causes of internal distractions were conversations with a passenger and use of tape decks. Today we have a host of modern telematic devices that offer even greater opportunities for internal distraction (Kantowitz, 2000).

The most familiar telematic device is the car radio. The Michigan legislators, after my prepared testimony, asked me several questions comparing radios to cell phones: e.g., we don't legislate any constraints on radios so why should we treat cell phones differently? But even the familiar radio is no longer your father's radio with one control for tuning and one control for volume. Figure 1 shows the percentage of car radios with less than eleven buttons over the last decade. While not the results of a random scientific survey, it clearly reveals that radios have become more complex, and hence more likely sources of internal distraction. For the five most recent

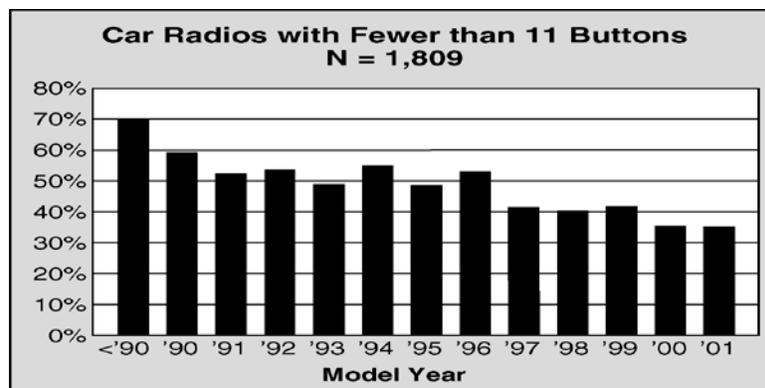


Figure 1 (AAA Foundation for Traffic Safety)

model years, more than half of installed car radios have eleven or more controls. Some of this complexity is good human factors, as when volume and seek controls are located on the steering

wheel, but most of this complexity has increased the driver's workload. So my answer to the legislators was that it depends upon the kind of radio.

Contemporary estimates of driver distraction are higher than those of the older tri-level study. A common lower bound for this is 26% based upon sampled crashes from the 1995 National Automotive Sampling System-Crashworthiness Data System that were attributed to driver inattention (Goodman et al, 1999). It is very difficult to determine to what degree specific devices or activities within the vehicle contribute to inattention and distraction; for example, food and beverages may be as important as cell phones (Hancock & Scallen, 1999). Case control studies that provide objective data are badly needed (Smiley, 1999). Yet it seems obvious that as cell phones and other telematic devices proliferate, increasing exposure to internal distraction will decrease driving safety. I do not believe it is prudent to wait until sufficient objective sources of data are available to start devising ways to mitigate driver distraction.

Two attorneys, highly familiar with human factors, have offered three risk reduction techniques intended to reduce driver distraction (Peters & Peters, 2001). First, telematic devices could have warning labels and messages. I doubt that this will be a solution to the problem. Second, is a binary integrated system with certain telematic devices being disabled while the vehicle is in motion. The notion of system integration is quite appealing, although this chapter will advocate a more continuous scheme for controlling in-vehicle devices using an intelligent interface. Third, is marketing and dealer restraint whereby risk reduction becomes a higher priority for the sales channel. This method of reducing risk is beyond the scope of the present chapter.

This chapter aims at exploring intelligent driver interfaces that minimize distraction from telematic devices. First, I define and explain my conception of an intelligent interface; these interfaces have the potential to minimize driver distraction. Then I distinguish between analytic and empirical approaches to evaluating an intelligent driver interface. My emphasis is upon empirical methods, especially those that utilize laboratory techniques. Discussion is focused upon microworlds, including driving simulators, that allow considerable laboratory control of experimental variables but still provide a reasonable approximation of the complexity of the real world. How might microworlds be used to evaluate driver interfaces? Since people drive vehicles on concrete highways, and not just in our constructed microworlds, what caveats and limitations need be taken into account before making practical recommendations to minimize driver distraction?

Defining Intelligent interfaces

Over a decade ago, I presented some thoughts about interfacing human and machine intelligence (Kantowitz, 1989). At that time I was judicious enough to deal only with qualitative aspects of intelligence:

Human intelligence is easy to define so long as one is prudent enough to refrain from attempts at measuring intelligence and is content to define intelligence through the behavior it produces. I define intelligent behavior as purposive behavior that attempts to reach a goal. For me this definition implies a closed-loop setting where feedback is used to modify behavior until the goal is attained (p. 50).

I then utilized this same definition for machine intelligence, making reference to Turing's test (Turing, 1950). This famous test calls for an observer to distinguish between a human and a machine. If the observer cannot do so, one must conclude that the machine is at least as intelligent as the human. Turing was able to anticipate potential objections to his test, including the objection that the intelligence resided in the programmer of the machine rather than the machine itself. Of course, if programmers always knew how their programs would behave it would never be necessary to debug programs. Many years ago, as a graduate student in computer science, I was required to write a SNOBOL program that played GO. My simple program was consistently able to beat me. (This, of course, is better explained by my inexperience as a GO player rather than attributed to any outstanding capabilities as a SNOBOL programmer.) Thus my GO program was more intelligent than I because it better achieved the goal of defeating its opponent.

If one admits that both human and machine system components can exhibit intelligence, the obvious question is how to link them to optimize overall system performance. This question usually is answered in human factors by referring to principles about allocation of function between people and machine (Kantowitz & Sorkin, 1987). In the traditional binary interface, an operator can manually decide whether to allocate some task to the machine or to perform it manually. Thus, if operator workload becomes too high more tasks can be assigned to the machine. But even this does not guarantee a sufficient decrease in operator workload since the operator must then keep track of the states of various sub-systems which itself increases operator workload.

How might one improve on the traditional binary interface? If we think of intelligent control of a system as a continuum, an optimal interface could assume any state along this continuum, **without creating any overhead cost** associated with either the state itself or the path used to reach the state. The ends of the continuum would be complete system control by either human or machine intelligence. While a binary interface can move along this continuum, although not always in a smooth graded manner, it creates overhead as it changes allocation of function. This overhead depends upon the number of machine sub-systems engaged and often also upon the order in which the operator addresses these sub-systems.

An optimal interface would degrade gracefully as workload increased. It would transfer tasks to the sub-system, human or machine, that would perform the best given their spare capacity and capability. For example, a task that could be performed with 90% efficiency for an unloaded human operator might be transferred to the machine, which only offered 72% efficiency when the human became overloaded enough so that his predicted performance dropped below 72%. If workload become excessive for both human and machine components, e.g., total system performance would be unsatisfactory, an optimal interface would shed load by deferring and even eliminating task components. Thus, an optimal interface would be adaptive. It would not have a fixed hierarchical list of tasks to be shed but instead would make these decision dynamically based upon real-time assessment and evaluation of sub-system resources.

Thus, an optimal interface must itself be intelligent. It must monitor operator performance and make allocation of function decisions for the operator. This represents a design philosophy that it much more acceptable in Europe than in the United States. For example, in Europe there is on-going research about controlling vehicle speed automatically by having controllers built into

every new engine. The machine would prevent the operator from speeding. It is hard to image such a system becoming popular in the United States even though it would greatly improve road safety. But I believe we have already exceeded the human driver's ability to safely control a vehicle while using all manner of telematic devices. For example, by 2002 electronic and electrical applications will account for 44 pounds of the 58 pounds of copper wiring in the average car. Much of this increase comes from telematics and entertainment. Intelligent interfaces must be devised.

An intelligent interface cannot be adaptive in an optimal manner without some knowledge, and perhaps even some preview, of the local environment. Monitoring the driver's workload is one important component of the local environment. There are many techniques for measuring workload (see Kantowitz & Campbell, 1996) but I have always liked using information theory (Kantowitz, 1985) and believe that one especially promising methodology that is practical in a moving vehicle is based upon the steering entropy (Boer, 2000). Unlike secondary-task and most physiological measures, steering entropy is unobtrusive. It is sensitive to demands of non-driving in-vehicle tasks. It can be related to models of attention, information processing, and closed-loop control. I hope more investigators incorporate this measure of performance into their research efforts.

Another kind of local knowledge relates to the driving environment outside the vehicle. For example, vehicles equipped with advanced cruise control already contain sensors that monitor distance and rate of closure to other vehicles. An intelligent interface could use this information to filter sources of in-vehicle information that are of lower priority. Similarly, existing road databases used with navigation systems could be extended to contain road accident data. An intelligent interface could use this information to limit in-vehicle distractions when approaching and traversing high-accident areas.

An intelligent interface capable of making dynamic allocation of function decisions for the driver must be designed carefully to prevent mode problems that have occurred in aircraft. Pilots have gotten confused about what the automation is doing due to insufficient feedback. Drivers must have a sound mental model that is consonant with the capabilities of vehicle automation. For example, it would be unsafe if drivers believed that a new adaptive cruise control totally removed the need for human monitoring of the vehicle because it would bring the vehicle to a safe stop should the preceding vehicle suddenly brake to a halt. While current adaptive cruise controls can slow the vehicle, they are not designed as safety devices that take the driver out of the loop. Any vehicle with more than a single intelligent controller, e.g., human and machine intelligence, must always have provision for strong annunciation of which intelligence is currently in control.

More recent developments have improved my qualitative meandering on intelligent interfaces by suggesting quantitative techniques to measure machine intelligence (Park, Kim & Lim, 2001) and interface complexity (Kang & Seong, 2001). For example, Park et al (2001) defined a control intelligence quotient by summing task intelligence costs across a task allocation matrix. They then perform a similar calculation for human intelligence based upon tasks allocated to the operator. The machine intelligence quotient is simply the control intelligence quotient minus the human intelligence quotient. As task allocation changes, it is easy to recalculate the effects upon both intelligent quotients. There is even a numerical example that illustrates these calculations.

My problem in fully understanding these concepts lies in measuring the task intelligence cost defined as a vector across tasks. The authors offer no advice on how to establish these numbers, other than suggesting one of six approved methods of measuring mental workload: parameters from behavioral signals, dual-task methods, information measures, eye scanning movements, subjective measurement, and physiological variables. I doubt that if any researcher was heroic enough to apply all six of these methods for the same interface, that they would all agree. Even more fundamental, the psychometric scale properties of these methods differ greatly. Although I am strongly in favor of computational models (Kantowitz, 2001), these models must be populated with measured quantities. Matrix multiplication of arbitrary numbers, or even numbers with unknown psychometric properties, can create only the illusion of precision. There is no psychological reality in the naked equation. While Park et al (2001) have provided an interesting discussion of how interface redesign through task allocation changes complexity and system intelligence, it is hard for me to accept that the numbers produced have even interval scale properties without knowing how the basic entries are measured. The outputs of even the most clever models cannot be better than the quality of the data entered into these models.

DESIGNING AND EVALUATING INTELLIGENT INTERFACES

Designing an intelligent interface presents many challenges. In particular, since new telematic devices are being developed and deployed rapidly, it seems difficult to anticipate all these new designs. One of the main challenges to the auto industry is to speed up its model cycle time to better match the rapid pace of the electronics industry that is creating new telematic devices. How can an engineer design an intelligent interface to accommodate devices that do not yet exist?

There are both analytic and empirical answers to this question. One analytic solution, using higher levels of abstraction to design based upon functions rather than upon specific devices, has already been articulated by Lee and Kantowitz (2001). It uses groups of functions and network analysis of information flows to integrate in-vehicle devices. Well before any physical device specifications are finalized, the design engineer knows what functions need be performed. This permits an evaluation of functions early in the design cycle. But even this useful analytic design tool does not remove the need for evaluating physical devices before they are deployed. This chapter focuses upon empirical evaluation and rapid prototyping by using microworlds to evaluate new telematic devices and their driver interfaces.

Microworlds and Driving Research

Human factors researchers must navigate a narrow course that weaves between the Scylla of lack of control in field studies and the Charybdis of sterile well-controlled laboratory studies that are difficult to generalize. As a graduate student I was trained in how to conduct conditioning experiments that were exquisitely controlled using the human eyeblink response. But I was unable to accept the fundamental principal of human eyelid conditioning: the human is appended to his eyelid. Given the extremely limited nature of this response measure compared to the larger repertoire of human thought and action, it was no surprise to me that human eyelid data exhibited much the same form and functional relations as found in rabbit eyelid studies. When a human is strapped into a chair, shown simple geometric forms on a screen, and has puffs of nitrogen directed onto his cornea, there is little opportunity for expressing a wide range of human

capabilities. But this research did engender a respect for experimental control that to this day makes me uncomfortable with some of the compromises often seen in field research.

Microworlds (Brehmer & Dörner, 1993) are computer-generated artificial environments that are complex (have a goal structure), dynamic (operate in real time) and opaque (the operator must make inferences about the system). Thus, they can avoid results of limited generality, e.g., some laboratory research on stimulus-response mappings is not helpful to human factors practitioners (Kantowitz, Triggs & Barnes, 1990), while maintaining a satisfactory degree of experimental control. Microworlds have been used to study topics such as process control (Howie & Vicente, 1998), extended spaceflight (Sauer, 2000), fighting forest fires (Omodei & Wearing, 1995), air traffic control (Bramwell, Bettin & Kantowitz, 1989), stock market trading and internet shopping (DiFonzo, Hantula & Bordia, 1998), and submarine warfare (Ehret, Gray & Kirschenbaum, 2000).

Ehret, Gray and Kirschenbaum (2000) have identified three useful dimensions that allow researchers to compare microworlds and other simulated task environments: tractability, realism, and engagement. Tractability relates to how the researcher can effectively use the simulated environment. For example, Bramwell, Bettin and Kantowitz (1989) used Seattle firefighters as subjects in their air traffic control microworld because the major aim of the study was to see how experienced teams with leaders of a known management style interacted. The goal was not to study air traffic control per se and the microworld created was an amalgam of enroute and terminal control. Two controllers (North and South) sat at different screens, which contained the same airport. In order for them to successfully route traffic, they had to coordinate their efforts because they could not see the other controller's screen. A team leader, who had no screen and thus was dependent upon the two controllers for information about traffic flow, was responsible for this coordination. This microworld exhibited tractability because it was simple enough for the controllers to learn quickly and easily how to command their aircraft, but still sufficiently complex to require teamwork and coordination. Realism refers to matching experience in the real and simulated worlds. For example, process-control microworlds should obey the same laws of thermodynamics as physical plants. Engagement refers to the willing suspension of disbelief on the part of experimental participants. Researchers want their subjects to act "naturally" and to produce the same behavior as in the real world. Engagement is a joint property of people and the simulated environment and the same microworld could be engaging for one person and not for another. Professionals, who are highly motivated and knowledgeable, might accept an unrealistic simulation because they can fill in missing details. For example, professional airline pilots will cheerfully fly a lower-fidelity Link trainer. Professionals might also reject a realistic simulation if it conflicts with their philosophical worldview. Military pilots might not wish to participate in studies of side-stick controllers because they believe that such devices do not belong in airplanes. I vividly recall one commercial truck driver who stormed out of our simulator after a crash because he had never experienced an accident in his entire career; did this mean he was too engaged?

These three concepts, tractability, realism, and engagement give us the vocabulary to compare microworlds with each other and with the real world. For example, why use microworlds to study driver distraction at all when that can be accomplished with greater realism and engagement in an actual vehicle? Safety is often cited as a major reason for using microworlds: no one ever got killed in a simulator crash. But we can safely use a real vehicle on a closed test

track to study driver distraction. It is fairly easy to instruct drivers to eat pizza or to insert CDs into a player while driving a closed course so that lane incursions and other vehicle control parameters can be recorded. One could even have plastic deer and foam-core cars unexpectedly pop out into the driver's lane. Indeed, these procedures are eminently suitable if the goal is to educate drivers about the dangers of distraction or to inform legislators about the need for regulating cell phones and perhaps other telematic devices. Videos of people driving real cars and crashing into obstacles while distracted are far more convincing than videos of simulated driving. No driving simulator has the pixel flow rate and transport lag of a real vehicle. But I believe that such test track work offers considerably less tractability when the goal is to improve the engineering design of in-vehicle interfaces. It takes much time and effort to conduct field research. While there are some savings in replications of field research, e.g., the same data acquisition system can be utilized, the virtual world offers far greater control and complexity. Complicated interactions between vehicles are easily programmed. Experimenters interested in control of braking systems on icy roads no longer have to wait for winter or travel to the far northern reaches to conduct their research. Drivers can be trusted to drive alone without an accompanying experimenter poised to apply an emergency brake should the vehicle run off the road. No tow trucks are required to restore such a vehicle to the roadway in the virtual world. Tractability is the reason for employing microworlds in driving research.

A Route Guidance Microworld

The Battelle Route Guidance Simulator (Kantowitz, Kantowitz, & Hanowski, 1995) consists of two linked Intel 486 computers and two video displays (Figure 2). One display shows real-time traffic video with rapid switching between traffic links and the other display provides a map with touch control selection of desired routes. The driver's goal is to reach a destination in minimum time. A moving dot indicates the vehicle's current location on the map. Using the touch screen, the driver can query traffic conditions on a particular link and select links for travel. The Route Guidance Simulator operates in real time on both screens so that a journey through the microworld takes exactly the same time as driving that route in real traffic. In order to increase engagement (and perhaps realism as well) a cash bonus, usually \$15, was continually displayed based upon travel time. Encountering traffic congestion severely depleted that bonus. A price could be charged for querying traffic links; in some experiments this price was zero. The accuracy of reported traffic congestion on a link was a major independent variable in these experiments. In selecting links, drivers were not permitted to choose a link beyond what was adjacent to the current link, i.e., drivers could not pre-program their entire journey. A link could not be selected unless it had first been queried for traffic congestion. Drivers proceeded through their journey selecting links and observing real-time traffic flow until they reached their destination.

This microworld exceeded our expectations for driver engagement. All sorts of emotional behaviors were observed as drivers fumed, stuck in heavy traffic –the queried reports were not always accurate- watching their bonuses evaporate. High levels of engagement need not be associated with positive feelings about the microworld. Indeed, only one observer, a traffic engineer who enjoyed watching congestion, was really happy with the simulated experience; this was fortunate since he was responsible for monitoring the project. Everyone else found the traffic delays to be irritating. While it was possible to change the time scale, e.g., running in half actual time and thus being twice as efficient in collecting data, we never did this because it

would have distorted the realism of the microworld. Similarly, we never found it necessary to increase the time scale making the simulation slower than real traffic but this might have been a useful technique for inducing road rage in the laboratory. Although we did not formally try to evaluate the realism of this microworld, it appeared that turning off the traffic video substantially reduced both realism and engagement.



Figure 2

The proof of the pudding is in the eating, and tractability is best evaluated by empirical results. Kantowitz, Hanowski, and Kantowitz (1997) used this microworld to investigate driver acceptance of unreliable traffic information in familiar (Figure 3) and unfamiliar (Figure 4) settings. Figure 3 is a map of Seattle. Figure 4 is a map of New City, a fictitious environment that no driver could have ever experienced. But look very closely at these two city maps. They are topographically identical with the same number and arrangement of nodes and links. (Rotating figure 4 90° makes this easier to see.) Only a microworld makes it easy to obtain this level of control. Comparing two real cities would confound familiarity and topography. Results of these experiments showed that drivers would accept unreliable information up to a point and that familiar driving environments require more reliable in-vehicle traffic information; see Kantowitz et al (1997) for detailed results and discussion of these points. I believe that the Route Guidance Simulator is a very effective research tool, especially when its low hardware cost is also considered.

The Route Guidance Simulator is a stand-alone part-task simulator. It can be used to discover design specifications for future in-vehicle devices. However, in order to use microworlds to investigate driver distraction a much larger microworld, consisting of a driving simulator and simulated telematic devices is required. The remainder of this chapter explores this larger virtual world.

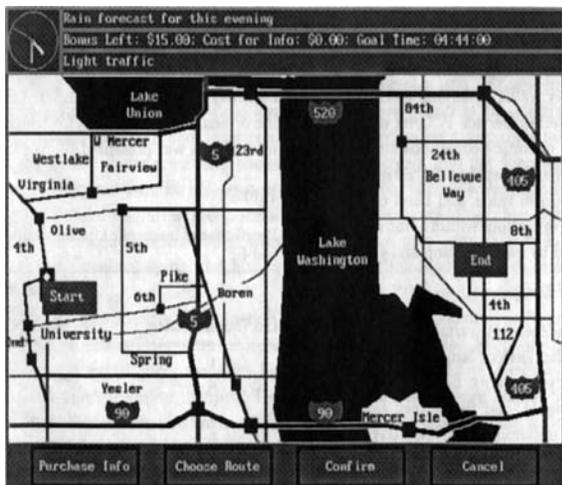


Figure 3

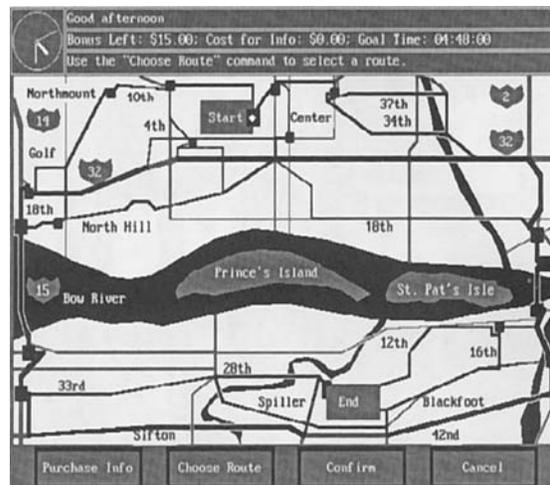


Figure 4

USING DRIVING SIMULATORS

Since no driving simulator can duplicate all the perceptual cues of the real world, it is important that the relevant cues be present with sufficient quality and quantity to allow generalization of results beyond the microworld. Successful microworlds need not require complete realism, as discussed previously. Indeed, simulator researchers have known for some time that psychological fidelity is as important as physical fidelity (Kantowitz, 1988; Rankin, 1984). The same driving simulator may be useful and valid for some applications and invalid for others. Thus, with each new application of a driving simulator, a new validation study should be conducted. Since those who fund research are usually more interested in obtaining results, rather than validating them, I commend any researcher who has been able to conduct validation studies of their simulator. It would be interesting to perform a literature review that compared the number of articles about how to build simulators and what a specific simulator will be able to do once completed to the number of articles that validated existing simulators by comparing them to real vehicles and to other simulators: in a perfect world validation studies would outnumber simulator construction articles by an order of magnitude.

A recent review of driving simulator validity (Kaptein, Theeuwes, & van der Horst, 1996) identified two important kinds of validity: absolute and relative. A simulator with absolute validity produces results and effect sizes that are identical to the real world. (Actually, this is my definition; Kaptein et al used the safer word "comparable" rather than identical, but I don't like hedging.) Relative validity means that treatment in the simulator produce the same rank order as in reality. They concluded that most simulator results produced relative validity, which is still worthwhile and useful.

Although this is a valuable distinction, I prefer to think of simulator validity more in terms of regression. How well does the simulator predict an outcome on the road? This allows for outcomes that are not absolute but still are better than relative validity. There is no need to settle for an ordinal scale when an interval scale might be achieved. Regression analysis also offers a metric that explains how well simulators predict reality so that different users can make their own judgments about the sufficiency of the fit for their own design purposes.

If one can obtain absolute validity, it is tempting to conclude that nothing else need be done. McGehee, Mazzae and Baldwin (2000) conducted a heroic validation study with 120 simulator drivers and 192 test track drivers. That's a lot of drivers! They found remarkable agreement and no statistical differences between simulator and test track total brake reaction time and time to initial steering. They were less fortunate in comparing time to throttle release but offered several plausible reasons to account for this statistically significant difference. Of course, since validity was absolute for two important measures, there was no need to offer plausible reasons for successful outcomes.

The strong implication of this approach is that equality of means produces validity. If means are unequal then results are not valid and plausible reasons must be invoked. I think this logic is incomplete on several grounds. First, why limit discussion to measures of central tendency? Distributions have been obtained and these should also be presented and compared. Furthermore, if reaction time measures are reported, mathematical psychology offers several techniques based upon distribution properties. Second, is the statistical problem associated with acceptance of the null hypothesis (Kluger & Tikochinsky, 2001). Using confidence intervals does not completely solve this difficulty. Instead, power analyses should be performed so that effect size can be determined (Cohen, 1988). I presume that McGehee et al (2000) prudently tested large numbers of drivers so that readers could not attribute the (desirable) lack of statistical significance to a weak experiment. A power analysis would confirm this presumption. If you have completed a powerful experiment, flaunt your results and show the power analysis. There are several computer programs that can do power calculations. Third, research is better guided by theory (Kantowitz, 2001). How should readers interpret the findings that maximum lateral acceleration was 1.24 g on the simulator and 1.17 g on the test track while maximum longitudinal acceleration was 0.65 on the test track and 0.8 g on the simulator? Without a model of the driver, it is difficult to answer the question. Fourth, the key question is not if simulator results equal test track results, although this is nice, but rather how well the simulator results predict the test track results. This is better examined with experimental designs that compare results over a set of parameters instead of obtaining only point estimates for a single event, such as a lane incursion. For example, driver speed could have been varied as a parameter.

Figure 5 shows hypothetical outcomes for such a parametric experiment. On-road values are plotted against simulator values. For example, one might compare curve entry speeds, for curves of different radius. Each curve would be represented as a single point in Figure 5. So curve radius would be the parameter that allows a state space to be constructed. If entry speed was identical for simulator and on-road tests, absolute validity, results in Figure 5a would be obtained. The squared correlation coefficient for this scatter plot would be an index of fit. But this is not the only favorable outcome. One might expect that because visual information flow is always lower in a simulator than on the road, drivers might consistently drive faster in a simulator. I would consider this also to be favorable outcome as shown in Figures 5b and 5c. On-road speed could still be predicted from the simulator and the correlation coefficient would indicate goodness of fit. Only if results of Figure 5d were obtained would I despair and conclude that a simulator is not a useful tool because knowing the simulator speed does not help in predicting the on-road speed. To recapitulate, the key to this kind of regression analysis is having some parameter that can be varied to sweep out the scatter plot.

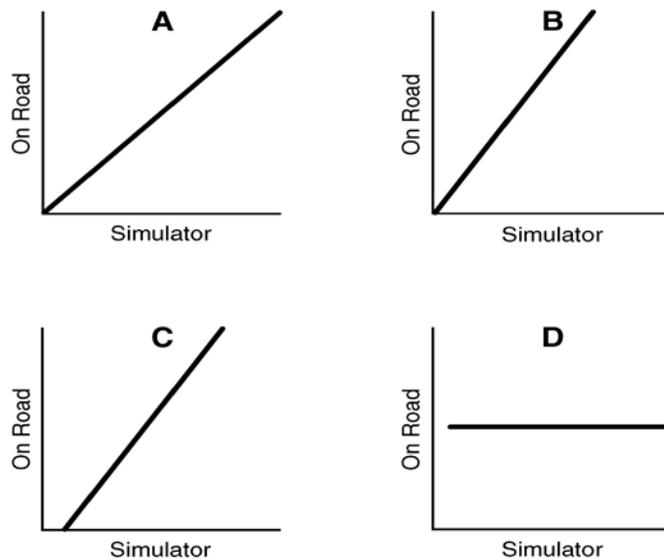


Figure 5

Bittner and Simsek (2000) presented such a scatter plot of curve-entry speeds and found a very low correlation ($R^2 = .06$) as shown in Figure 5d and so concluded that the simulator was not a valid tool for this purpose. Their description of the experiment was terse and readers could be forgiven for believing that the simulated and on-road courses were identical. However, this was not the case. The simulator track contained orthogonal sets of curves with varying degrees of radius and deflection, an orderly environment that could be established only in a microworld. A small number of curves that matched a real road in southern Washington state was sprinkled among these orthogonal curves. Each curve was preceded by a tangent (straight section of road). The curve preceding each tangent was in no way matched to the real road. Since in many cases the speed on the tangent was limited by its preceding curve, I would not expect correspondence in matched curve-entry speed unless speed on the tangent happened to be identical in simulator and road experiments. A better way to analyze these data, therefore, would be correct for speed on the tangent before comparing curve entry speeds. When this is done, results look like Figure 5c. The correlation is quite high using adjusted speeds. In this experiment, the same drivers drove the simulator once and the real road twice. It would be unreasonable to expect the simulator to correlate with the road more highly than the road correlates with itself. So the experimental design allowed correction for attenuation based upon repeated measurements on the road. This produced an extremely high correlation, $R^2 > .90$.

Of course, the best way to design a validation study would be to have the simulator roads be an exact copy of the real road. However, this requires a sponsor willing to pay for a pure validation study. The study just described was a compromise aimed at examining effects of road geometry with just a little bit of validation on the side. Nevertheless, when analyzed correctly, it revealed that the simulator is an excellent tool for estimating curve entry speeds. Although the fixed-base simulator did not provide lateral acceleration cues, which probably would be necessary for studying control of speed through a curve, these acceleration cues are not important on tangents. The orthogonal microworld allowed me and my colleagues at Battelle, where the study was conducted, to discover that curve entry speed was influenced by radius of curvature (no surprise here) and also by deflection angle (big surprise to highway engineers). I doubt that this question could have been investigated by using existing highways. First, it would be almost impossible to

locate a sufficient number of curves with appropriate combinations of radius and deflection angle within reasonable driving distance. Second, many other road geometry parameters would have confounded results so that the effects of deflection angle might not have emerged. The tractability of the real world for this experiment is close to zero. Yet the experimental outcome from the microworld will be of great interest to highway engineers who want to design self-regulating highways where drivers are guided to control vehicle speed by their perceptions of the roadway rather than by road advisory signs.

MICROWORLDS AND DRIVING DISTRACTION

The preceding discussion has explained some properties of intelligent interfaces, microworlds and driving simulators. It is now time to assemble these pieces, although many readers have probably already extrapolated to these conclusions. But first a comment on the current intense focus of researchers and politicians on driver distraction. The term “driver distraction” can be a bit prejudicial since it implies that some irrelevant sources of information should be eliminated. The world needs a technological solution to purge distraction and if that cannot be provided then perhaps a legislative solution will emerge. I think this is putting the cart before the horse. My preference is to apply good engineering design to in-vehicle information systems. I believe that if this is done using best ergonomic practices, then the issue of additional driver distraction created by telematics will be moot. Our goal should be to improve the driver interface. If this is accomplished then telematics will increase driving safety and distraction will be limited to social interactions within the vehicle, grooming, mastication and other dalliances unrelated to in-vehicle technology. Telematics should never be a source of driver distraction.

This happy state will be achieved most rapidly by using microworlds to evaluate the design of intelligent interfaces that prevent driver distraction by controlling the flow of information and control within the vehicle. While new analytic tools may assist in narrowing down possible design alternatives, final designs must be evaluated before being made available to the driving public. Virtual tools are the best way to accomplish this goal. It is not surprising that the first generation of telematic devices can distract drivers. These new systems work by themselves with little or no integration. They can present conflicting and confusing information. But this can be cured with the same kind of large system integration already found in airplanes and process control plants. I believe that using microworlds will hasten the integration of in-vehicle telematics.

ACKNOWLEDGMENT

This research was supported by the UMTRI Director’s fund for improving driving science. Any correct statements in the article are solely the responsibility of the author and have not been endorsed by the Regents of the University of Michigan.

REFERENCES

- Bittner, A. C., & Simsek, O. (2000) Simulator validation: Methods and illustrations. *ISATA International Symposium on Automotive Technology and Automation, 33rd, Simulation and Virtual Reality* 63-69. Epsom, England, ISATA.
- Boer, E. R. (2000) Behavioral entropy as an index of workload. *Proceedings of the IEA/HFES 2000 Congress, 3*, 125-128.
- Bramwell, A., Bettin, P., & Kantowitz, B. H. (1989) The effect of leadership style on team performance of a simulated air traffic control task. *Paper presented at Western Psychological Association*.
- Brehmer, B., & Dörner, D. (1993) Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of field study. *Computers in Human Behavior, 9*, 171-184.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. 2nd ed., Mahwah, NJ: Erlbaum.
- DiFonzo, N., Hantula, D. A., & Bordia, P. (1998) Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods, Instruments, & Computers, 30* (2), 278-286.
- Ehret, B. D., Gray, W. D., & Kirschbaum, S. S. (2000) Contending with complexity: Developing and using a scaled world in applied cognitive research. *Human Factors, 42* (1), 8-23.
- Goodman, M. J., Tijerina, L., Bents, F. D., & Wierwille, W. (1999) Using cellular telephones in vehicles: Safe or unsafe? *Transportation Human Factors, 1* (1), 3-42.
- Hancock, P. A., & Scallen, S. F. (1999) The driving question. *Transportation Human Factors, 1* (1), 47-55.
- Howie, D. E., & Vincente, K. J. (1998) Measures of operator performance in complex, dynamic microworlds: Advancing the state of the art. *Ergonomics, 41* (4), 485-500.
- Kang, H. G., & Seong, P. H. (2001) Information theoretic approach to man-machine interface complexity evaluation. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans, 31* (3) 163-171.
- Kantowitz, B. H. (2001). In-vehicle information systems: Premises, promises, and pitfalls. *Transportation Human Factors*. (In press).
- Kantowitz, B.H. (2000). Effective utilization of in-vehicle information: Integrating attractions and distractions. *Proceedings of Convergence 2000 International Congress on Transportation Electronics*, 43-49. Detroit, MI: Society of Automotive Engineers.
- Kantowitz, B. H. (1989). Interfacing human and machine intelligence. In P. A. Hancock & M.H. Chignell (Ed.), *Intelligent Interfaces: Theory, Research and Design*. 3, 49-67. North-Holland.

Kantowitz, B. H. (1988). Laboratory simulation of maintenance activity. *Proceedings of the 1988 IEEE 4th Conference on Human Factors and Nuclear Power Plants*, 403-409. New York: IEEE.

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

Kantowitz, B. H., & Campbell, J. L. (1996). Pilot workload and flightdeck automation. In R. Parasuraman & M. Mouloua (Eds.), *Human performance in automated systems* 117-136. Mahwah, NJ: Lawrence Erlbaum Associates.

Kantowitz, B. H., & Sorkin, R. D. (1987). Allocation of functions. In G. Salvendy (Ed.), *Handbook of human factors*. New York: Wiley.

Kantowitz, B. H., Hanowski, R. J., & Kantowitz, S. C. (1997). Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors*, 39(2), 164-176.

Kantowitz, B. H., Triggs, T. J., & Barnes, V. (1990). Stimulus-response compatibility and human factors. In R. W. Proctor & T. Reeves (Eds.), *Stimulus-response compatibility*, 365-388. Amsterdam, The Netherlands: North-Holland.

Kantowitz, S. C., Kantowitz, B. H., & Hanowski, R. J. (1995). The battelle route guidance simulator: A low-cost tool for studying driver response to advanced navigation systems. In D. J. Dailey & M. P. Haselkorn (Eds.), *Proceedings of the Sixth Annual International Conference on Vehicular Navigation and Information Systems (VNIS'95)*, 104-109. Piscataway, NJ: IEEE.

Kaptein, N. A., Theeuwes, J., & van der Horst, R. (1996) Driving simulator validity: Some considerations. *Transportation Research Record 1550*, 30-36 National Academy Press, Washington, D.C.

Kluger, A. N., & Tikochinsky, J. (2001) The error of accepting the "theoretical" null hypothesis: The rise, fall, and resurrection of commonsense hypotheses in psychology. *Psychological Bulletin*, 127 (3), 408-423.

Lee, J. D., & Kantowitz, B. H. (2001) Network analysis of information flows to integrate in-vehicle information systems. Working paper.

McGehee, D. V., Mazzae, E. N., & Baldwin, G. H. S. (2000) Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track. *Proceeding of the IEA 2000/HFES 2000 Congress*, 3, 320-323.

Omodei, M. M., & Wearing, A. J. (1995) The fire chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers*, 27 (3), 3003-316.

Park, H.-J., Kim, B. K., & Lim, K. Y. (2001) Measuring the machine intelligence quotient (MIQ) of human-machine cooperative systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31 (2), 89-96.

Peters, G. A., & Peters, B. J. (2001) The distracted driver. *Journal of the Royal Society for the Promotion of Health*, 121 (1), 23-28.

Rankin, W. L. (1984) Nuclear power plant simulators for operator licensing and training. *U.S. Nuclear Regulatory Commission*. BUR/CR-3725. Washington D.C.

Sauer, J. (2000) The use of micro-worlds for human factors research in extended spaceflight. *Acta Astronautica*, 46 (1), 37-45.

Smiley, A. (1999) Using cellular telephones in vehicles: safe or unsafe? *Transportation Human Factors*, 1 (1), 57-59.

Treat, John R. (1980). A study of precrash factors involved in traffic accidents. *HSRI Research Review*, 10 (6) 1-35.

Turing, A. M. (1950) Computing machinery and intelligence. *Mind*, 59, 433-460.