ADDRESSING SITUATION AWARENESS

While much of this discussion piece has focused on surrogates that are event-based, one finding from Angell (Chapter 3) calls for further discussion. It relates to the fact that subjective measures of situation awareness reported in the work by Young, et al. (2005) performed relatively well in the effort to establish criterion validity (meaning that it correlated well with on-road measures of situation awareness). In the Young, et al. (2005) work, the ratings were based on an adaptation of a NASA-TLX scale. This result has sometimes been greeted with surprise when there has been an expectation that objective measures would be more easily predicted than subjective rating measures. The fact that it seems surprising to many in the field is something that prompts discussion here, in the context of another finding. In the Young and Angell (2003) work, ratings of situation awareness tended to be associated with "amount of workload," loading on the first Principal Component that emerged. This is sensible, and interesting.

It suggests perhaps two things. First, there appears to be a role that conscious awareness of the situation plays in the process of scanning the roadway, noticing what's on it, attending to it, and responding to it. Second, there may be an especially good opportunity at this point in time for the considerable work on Situation Awareness (see Endsley and Garland, 2000; and Bolstad, et al., Chapter 11 in this book) to make contributions of constructs and methods toward a broader understanding in this domain, one that will integrate with it the understanding that has been emerging from event detection-and-response methodologies.

CONCLUSION

Comparisons of the Modified Sternberg Method to Peripheral Detection Tasks (of the type implemented in the CAMP DWM Study) suggest that it currently offers the strongest criterion validity across the broadest set of conditions (task types and locations of events-to-be-detected) for on-road event detection (to the extent that it has been measured to date). More importantly, however, exploratory analyses of the Sternberg method, of these particular Peripheral Detection Task methods, and of the Lane Change Task, illustrate the direction that future work could take – in confirming what these surrogates are measuring, in enhancing and perhaps combining these techniques, and in understanding how to modify or augment task design to improve situation awareness and event responsiveness. A number of issues were identified, which warrant further investigation. Finally, the possibility that the body of work in Situation Awareness Theory and Method could contribute to progress in this area was identified as promising.

REFERENCES

1. Angell, L. S. (2010). Conceptualizing Effects of Secondary Task Demands on Event Detection During Driving: Surrogat Methods and Issues. Chapter 3 (this book), SAE International, Warrendale, PA.
2. Angell, L., Auflick, J., Austria, P. A., Kocchar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J. Kiger, S. (2006). *Driver Workload Metrics Task 2 Final Report & Appendices*. Final Research Report: DOT HS 810 635. NHTSA/USDOT, Washington, D.C. Available at NHTSA Website http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf
3. Benedict, D.R., & Angell, L.S. (2010). Modified Sternberg Method for assessing event detection while driving. Chapter 6 (this book), SAE International, Warrendale, PA.
4. Benedict, D., Angell, L. & Diptiman, T. ***Exploration of the Lane Change Test. (2006)*** Presentation for the Driver Metrics Workshop sponsored by the Alliance of Automotive Manufacturers, SAE, and ISO. Ottawa , Canada.
5. Bolstad, C.A., Cuevas, H. M., Wang-Costello, J.J., Endsley, M.R., & Angell, L. S. (2010). Measurement of situation awareness for automobile technologies of the future. Chapter 11 (this book), SAE International, Warrendale, PA.
6. Borowsky, A., Shinar, D., and Parmet, Y. (2008). The relation between driving experience and recognition of road signs relative to their locations. ***Human Factors***, ***50***, 2, 173-193.
7. Endsley, M.R., & Garland, D. J. (2000). ***Situation awareness analysis and measurement***. Mahwah, New Jersey: Lawrence Erlbaum Associates.
8. Engström, J., Åberg, N., Johansson, E., & Hammarblick, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment and in-vehicle information systems (IVIS). In Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Rockport, Maine.
9. Engström, J., (2010). The Tactile Detection Task as a method for assessing drivers' cognitive load. Chapter 5 (this book), SAE International, Warrendale, PA.
10. Mattes, S. (2003). The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In: Strasser, H.; Kluth, K.; Rausch, H.; Bubb, H. (Eds.), ***Quality of Work and Products in Enterprises of the Future*** (pp. 57-60). Stuttgart: Ergonomia Verlag.
11. Mattes, S., Föhl, U., & Schindhelm, R. (2007). Empirical comparison of methods for off-line workload measurement. AIDE Deliverable 2.2.7, EU Project 1ST-1-507674-IP.
12. Mattes, S., Föhl, U., & Strus, S. (2004). Cognitive Calibration Task. Daimler-Chrysler Presentation. Germany.
13. Mattes, S. & Hallen, A. (2008). Surrogate Distraction Measurement Techniques: The Lane Change Test. In Regan, M.A., Lee, J.D., Young, K. (Eds.), ***Driver Distraction: Theory, Effects, and Mitigation*** (pp. 107-121). London: Taylor & Francis.
14. Sternberg, S. High-speed scanning in human memory. ***Science***, 1966, ***153***, 652-654.
15. Van der Horst, R. A. & Martens, M. (2010). The Peripheral Detection Task (PDT): On-line Measurement of Driver Cognitive Workload and Selective Attention. Chapter 4 (this book), SAE International, Warrendale, PA.
16. Van Winsum, W., Martens, M. & Herland, L. (1999). The Effect of Speech versus Tactile Driver Support Messages on Workload, Driver Behavior and User Acceptance. TNO-Report TM-99-C043, Soesterberg, Netherlands.
17. Wickens, C. J. (2002). Multiple Resources and Performance Prediction. Theoretical Issues in Ergonomic Sciences, 3 (2), 159-177.
18. Young, R. A. & Angell, L. S. (2003). The dimensions of driver performance during secondary manual tasks. Proceedings of "Driver Assessment 2003: The Second International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design", Park City, Utah. July 2003.
19. Young, R.A., Aryal, B.J., Muresan, M., Ding, X., Oja, S., Simpson, S. N. (2005). "Road-to-lab: Validation of the static load test for predicting on-road driving performance while using advanced in-vehicle information and communication devices," *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Paper available at: http://sun3.science.wayne.edu/~mttc/Articles/Website-Publications/young_2005.pdf

*Linda S. Angell was affiliated with General Motors at the time of this work. She is now affiliated with Virginia Tech Transportation Institute as a visiting Research Scientist.

ENDNOTES

1 The method used by van der Horst & Martens (Chapter 4) is focused on assessing the uniquely cognitive or attentional issues involved in performance. This is distinct from methods which are intended as a type of emulation of on-road event detection. In the latter, event detection on the road involves looking at the road scene, visually scanning it to detect events that require attention, in addition to the processes of attending to and responding to events – all of these processes are captured in the way the event-detection-emulation PDT methods are implemented (as described here). In contrast, the van der Horst & Martens (Chapter 4) focus on cognitive processes is achieved through the use of a head-mounted light-presentation methodology, which eliminates the effects of visual scanning processes on light detection and allows isolation of cognitive processes (because the light can always be seen whenever it is illuminated.) Variants of detection-tasks which use other (non-visual) modalities are intended for the same purpose as the van der Horst & Martens technique (see Engström, et al., 2005, and Engström, Chapter 5 in this book) – to isolate effects on cognitive processes.

2 The PDT lights in this method had fixed-duration exposures of 1-second, rather than being terminated by a subject response – as is done in the "standard" PDT (cf. Van Winsum, Martens, & Herland (1999)).

3 The Modified Sternberg Task was adapted from the original Sternberg method (Sternberg, 1966). In it, a participant is given a "memory set" of items to remember (usually numbers), and is asked to commit them to memory in advance of a set of test trials. A participant is then shown a series of items, one-at-a-time. When each item appears, the participant is asked to indicate (using a button press) whether it was -- or was not -- in the memorized set. The Sternberg method was adapted and applied as a laboratory surrogate method for the Driver Workload Metrics project (Angell, et al., 2006), as a way to identify tasks that not only imposed visual workload, but also imposed workload on working memory. Instead of using numbers as the items to be remembered in the memory set, it utilized stimuli which were related to the driving task – namely, road signs of two types. One type consisted of highway marker signs, which displayed numerals (and could be called verbal stimuli, since they can be recoded verbally for memory rehearsal). The other type of road signs were intersection warning signs (spatial stimuli which consisted of representations of intersections and roadway junctions).

Samples of these are shown below. In the Modified Sternberg Task, the probes from the memory set were shown every 2 to 10 seconds while a participant was simultaneously performing a secondary task (in order to evaluate the effects of the secondary task's load on performance). A fixed memory set size of 3 items was used. Stimulus types were not mixed within a trial, but alternated between trials.

Sample Stimuli for the Modified Sternberg Task (from Benedict and Angell, Chapter 6)

The Modified Sternberg Method generates a number of metrics, but those that proved most useful in the CAMP Driver Workload Metrics study (Angell, et al., 2006) were % Missed Detection, % Incorrect Responses, Given A Correct Detection, % Overall Error, and Combined Dual Task Decrement Score. See Benedict & Angell (Chapter 6) for a definition of each of these metrics.

4 As a research practicality, the on-road event detection paradigms used in the CAMP DWM study are used as if they are "ground truth" measures for actual event response performance during natural driving. However, it is still unknown to what degree these on-road event paradigms elicit the same kinds of driver behavior as naturally-occurring events. Discovering to what degree this may be true is a very difficult research challenge, for naturalistic and experimental approaches alike. The most that can be done at the present time is to explicitly identify it as an area in need of further examination.

5 The value of r = 0.707 was used as a criterion in the CAMP DWM study, based on a decision by the CAMP DWM research team that any useful surrogate metric should account for at least 50% of the variance in a corresponding variable associated with on-road behavior. A correlation coefficient of r = 0.707, when squared, corresponds to an R^2 value of 0.50 (with rounding) (or 50% of the variance). Thus, only correlations above this value are reported in the CAMP DWM Final Report. In addition, a criterion value set at this level assured that only correlations with statistical significance were reported – since, for example, with only 7 visual-manual tasks tested in the on-road venue, a correlation had to be greater than or equal to 0.67 to reach significance at $p < .05$.

6 Task descriptions for each of the tasks which could be compared are provided below. Further details can be found in the CAMP Driver Workload Metrics Final Report, Appendix B (Angell, et al., 2006). The tasks were:

Just Driving: This task was a two-minute task used to create a baseline for driving. It consisted of having the participant "just drive" and respond to any event-detection stimuli that occurred – but was free from any multitasking of secondary task activity. It was "just driving".

Radio Tuning (Hard): The radio hard task involved turning on the radio, changing band, and tuning to a specific frequency, which was near the end of the frequency band and quite far from the start-state of the task. The target frequencies were 35-37 increments of .0.2 MHz up or down from the start-state frequency to which the radio was set at the beginning of the task.

Phone (Manual Dial of Handheld Phone): The participant was asked to use a handheld cell phone, which had been positioned in the flipped-open position, to dial their home phone number (including area code), pressing "end" instead of "send" (to prevent the home phone from ringing repeatedly throughout the experiment).

Map Search (Hard): The participant was asked to give the relative orientation of two locations on a map. The map was a paper map, printed in color, with black callouts marking the locations. In the "hard" version of the task, there were 22 locations marked on the map – so the participant was searching for 2 among the 22 locations, in order to report their orientations relative to each other (e.g., Location 1 is northwest of Location 2).

7 "Task demand" is a construct which is intended to capture the totality of resources required of the human operator to perform a task successfully. Comprehended within this construct are three key elements: the **types of resources** that are required to successfully perform a task (e.g., visual, manual, auditory, vocal, cognitive, etc., -- see, for example, Wickens, 2002), the **amount of resources** of each type required, and the **pattern of resource requirements across time** during the period of task performance. A secondary task, for example, might demand a great deal of visual resources, some manual resources, and virtually no cognitive resources in order to be performed – and the visual demand may be intermittent across the task period, while the manual demands fall primarily at the end of the task period. Another task may have a completely different task demand profile.

CHAPTER 8

Naturalistic Driving: Crash and Near Crash Surrogate Distraction Metrics

Miguel A. Perez, Linda S. Angell, and Jonathan M. Hankey
Virginia Tech Transportation Institute

INTRODUCTION

A substantial amount of recent and ongoing work utilizes the naturalistic driving paradigm, with different interpretations in terms of the length of time and circumstances under which driver behaviors are observed. In this chapter, naturalistic driving research refers to the extended observation of driver behaviors, exposures, and actions occurring in real-world driving. The observation period may range from weeks to one or more years and occur in the participant's own vehicle or on a fleet of research vehicles, depending on the goals of the study. If there is an intervention present it is usually not highlighted to the driver, who is allowed to adapt to it and use it at his own pace and for a reasonable amount of time. There is no experimenter present in the vehicle, and, more often than not, the driver controls the timing, length, duration, and type of trips that are taken.

Naturalistic observation of drivers has been used as a research approach for a number of years. For example, much of the work on intersection approach behavior (e.g., Horst and Wilming, 1986), seat belt usage (Krafft, Kullgren, Lie, and Tingvall, 2006), and driver error (Wierwille, Kieliszewski, Hanowski, Keisler, and Olsen, 2002), has relied on observation of drivers as they perform different driving maneuvers. These observations typically take place at a roadside location. The main limitation of this approach has been, in many cases, the lack of context for the driver actions that are observed. Driver actions and behavior prior to the observation, which can be important modifiers of the types of errors that drivers make, could not be accurately detected.

As data collection technology has progressed, it has allowed the naturalistic observation process to take place within the confines of the vehicle. While in previous efforts this type of data collection effort was limited in duration and in the scope of the variables collected, the 100-Car Study (Dingus, Klauer, Neale, Petersen, Lee, Sudweeks, et al., 2006) produced a naturalistic driving dataset that overcame many of these past limitations. Other subsequent studies have used a similar research approach in a Field-Operational-Test research paradigm (LeBlanc, et al., 2006; University of Michigan Transportation Research Institute & General Motors, 2005). In these cases, the main goal has been the real-world pre-deployment evaluation of one or more crash countermeasure suites.

Naturalistic driving research cannot answer all questions about driving behavior, but it has provided important insight into how crashes happen. For example, the principal focus of the 100-Car Study was to understand the causal factors for different crash types and quantify how often these factors were present in everyday driving tasks. A number of secondary goals, however, were also established. These secondary goals included the characterization of driver inattention and the determination of presence and extent of driver behavior changes due to the presence of different devices and vehicle instruments. Potential uses of 100-Car data, however, extend well beyond this set of original data analysis goals. For example, the dataset may be used to aid in the development of driver models that describe following behaviors, response patterns, lane-change sequences, intersection crossings, and use of in-vehicle systems, to name a few. Many or all of these models depend on driver behaviors and performance (i.e., driver metrics) that may be affected by secondary or tertiary task performance, and thereby may be used as measures of the extent that these activities may be taxing to the driving task.

This chapter describes a number of alternative applications for naturalistic driving data such as that obtained for the 100-Car Study, all related to the development and measurement of driver metrics and surrogate measures of safety. Before describing these applications, a framework for the required instrumentation and available data is presented. Information on crash causal factors observed as part of the 100-Car Study is also presented. The fundamental problems of interest in this research area are described and different applications of naturalistic data in the realm of driver modeling are then discussed.

OVERVIEW OF A NATURALISTIC STUDY

Setting up the instrumentation and logistics for a naturalistic study is not a trivial endeavor. This section briefly describes typical approaches used in the past to successfully conduct one of these studies, with emphasis on those used for the 100-Car Study and other subsequent similar efforts at the Virginia Tech Transportation Institute.

INSTRUMENTATION

Naturalistic driving instrumentation packages are designed to operate continuously once a vehicle's ignition is turned on (and after a system initialization period has elapsed) and until the driver turns the ignition off. Systems are typically designed around a computer processor running custom data acquisition software and communicating with the different data acquisition nodes through a distributed network. (Figure 1 shows an example for 100-Car). Data collection systems have to be designed to be rugged, durable, maintainable, and unobtrusive. Cameras to collect video data are typically small and mounted in unobtrusive locations. Figure 2 shows a sample of typical video. Typical video images used include the forward scene, driver's face, a view of the instrument cluster and/or center console over the driver's shoulder, and a rear and/or side view(s). Data collection rates are application-dependent, but are typically 30 Hz for video data and 10 Hz for other data streams.

SUBJECTS

Participant demographics and the number of participants also have to be carefully considered to be compatible with the goals of the study. While studies of normative driving may attempt to employ a diverse range of participant demographics, studies that focus in particular technologies or features may concentrate on drivers that are more likely to use the technology or feature in the real world. In some cases, it may also be important to screen drivers based on their driving habits (e.g. to avoid spending project resources on drivers that do not drive enough to provide sufficient data). Participant demographics may also be related to the vehicle-type that is driven. Particular vehicles (e.g. luxury) that may be of interest in the study may be more likely to be driven by particular driver populations.

VEHICLES

Vehicles are typically included in a naturalistic study in one of two ways. First, it may be desirable to use the participant's own vehicle. In that case, the vehicle is typically taken from the participant for a brief period of time, instrumented, and returned for normal use. In many cases, the data needed also constrain the number of models, model years, and makes than can be instrumented. Second, and especially when particular technologies are being evaluated, it may be necessary to have participants use a pre-instrumented vehicle that is provided to them at no cost. In those cases the participant is asked to replace their personal vehicle with the instrumented vehicle for the duration of the study.

DATA COLLECTION AND REDUCTION

Data retrievals are typically scheduled throughout the week when the vehicles are not likely to be used, with minimal or no interaction between participants and experimenters. Retrievals may require access to the vehicle or trunk (access which is ensured during the participant introduction to the study), or may be wireless. The data obtained from the vehicles are then typically post-processed to ensure quality, detect correctable failures in data collection equipment, and to identify events of interest. Depending on the goals of the study, these events may be reviewed by trained data reductionists, who may determine the validity of the event, the associated severity, and a large number of other event categorization and identification variables, for example. In some studies event detection may occur based on observation of the complete video, or perhaps randomly selected segments. Eye-glance reduction around the event is also performed frequently on these types of datasets. The resultant reduction is usually stored along with the data in a secured database for use in future research endeavors. More information about the event identification and reduction process can be found in Dingus, et al. (2006).

DATA ACQUISITION SYSTEM FLEXIBILITY

While Figure 1 illustrates the DAS components that were used for the 100-Car system; it is important to note that each of these nodes is both independent from other nodes and only represent a sample of the nodes that can be made available for data collection efforts. For crash-modeling purposes, for example, it might be desirable to increase the accuracy of the accelerometer and its data collection rate. Other project needs may require different video views, links to other vehicle network variables, or direct links to other original or add-on sensors within the vehicle.

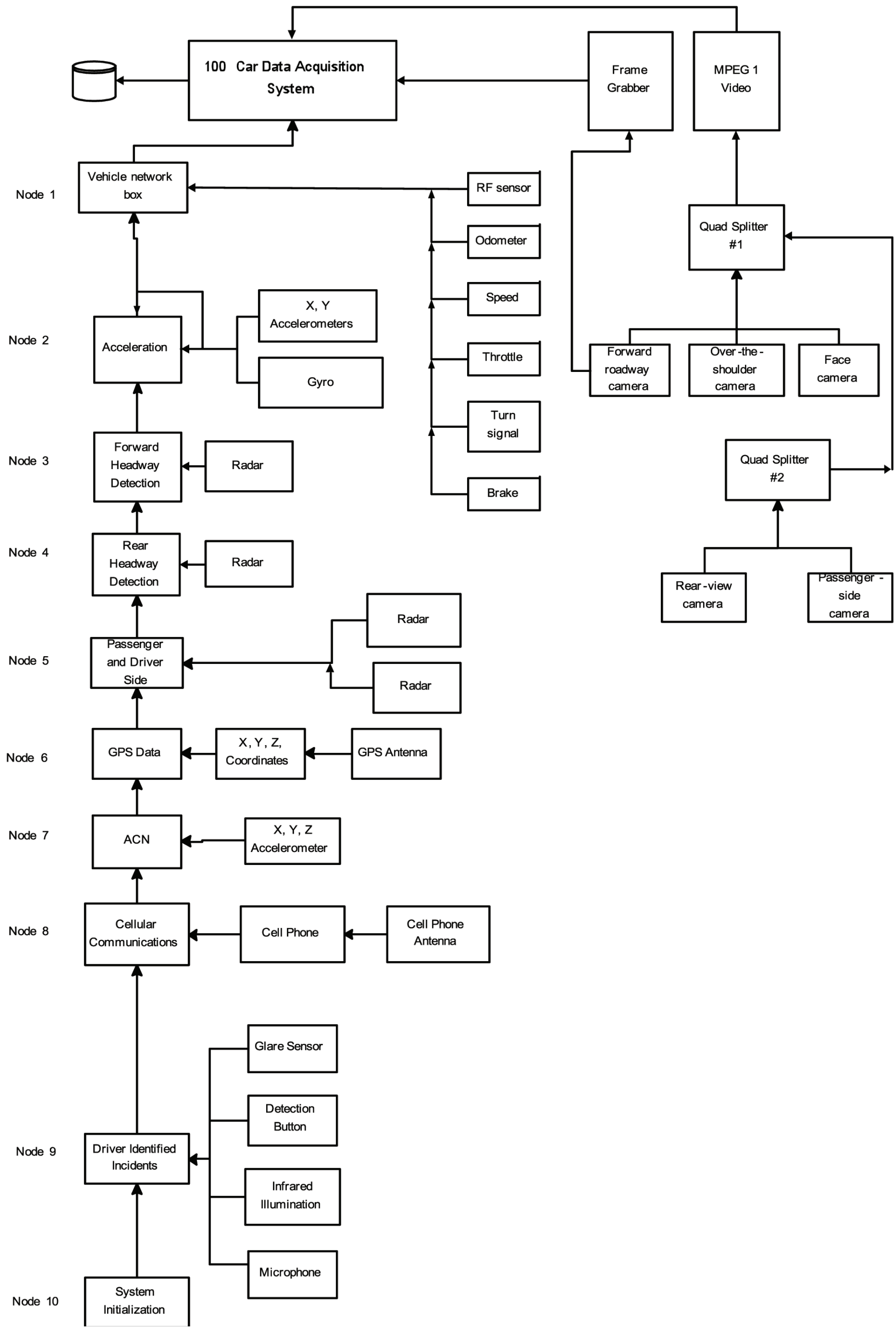

Figure 1: 100-Car data hardware collection system schematic

Figure 2: A video image from the 100-Car data. The upper left quadrant contains the view of the driver's face, which has been distorted to protect the driver's identity.

SUMMARY OF 100-CAR STUDY RESULTS

The uniqueness of the 100-Car Study lays in the 82 crashes and 761 near-crashes that participant drivers were involved in during a 1-year period. Crashes and near-crashes were defined to consist of driving events which differed only based on the success of an avoidance maneuver (i.e., unsuccessful or inexistent avoidance maneuvers result in crashes). These driving events were assumed to have similar kinematic and behavioral characteristics prior to the avoidance maneuver, an assumption that has been empirically confirmed (Dingus, et al., 2006; Klauer, Sudweeks, Hickman, and Neale, 2006). The availability of near-crash events yields a much larger and richer set of data than would have been attainable using epidemiological crash data alone, in that those same near-crashes would have simply never shown up in epidemiological studies or police/insurance crash reports, which typically concentrate on more severe events. Every near-crash event also demonstrates a driver successfully performing an evasive maneuver. This may provide additional insight into where and how human failure occurs during a crash.

Data from the 100-Car study provided a unique understanding on the prevalence and importance of several crash causal factors. A detailed discussion of these results is available in Dingus, et al. (2006); some highlights include:

- Driver drowsiness is involved in over 4 times the crash and near-crash events than previously thought (accounting for ~20% of crash and near crash events).
- Visual inattention to the forward roadway was a primary causal factor in the majority of crash and near-crash events (compared to 25-30% suggested by crash statistics).
- Interestingly, following a vehicle too closely produced a type of "protective effect" relative to normal driving. That is, drivers were typically found to be very vigilant when they chose shorter headways. In contrast, when drivers adopted longer headways *for the purpose of engaging in a secondary task*, they tended to under-compensate, resulting in a crash rate about 2 times greater than normal driving.

This last point is particularly interesting in that it goes against the popular belief that the cause of most rear-end crashes is tailgating. These results suggest that tailgating can be a factor, but it is by no means the only factor in rear-end crashes, at least in the population of drivers studied. Findings such as these can and should influence our understanding of surrogate measures of safety and driver metrics, and showcase how naturalistic observation can yield new and meaningful insights into real-world driving behavior.

APPLICATIONS OF NATURALISTIC DATA

SCOPES OF NATURALISTIC DATA ANALYSIS

Understanding the causal factors for crashes is a complex endeavor, moderated to some extent by the level at which that understanding is desired. At a macro or aggregate level (e.g. crash databases), there are robust data as to what the primary crash types are, their prevalence, and some indication of their causes. Of course, these causes can be the source of argument, as it is often the case that these databases are riddled with unknowns for these categories. The data they provide, however, is very important in that it frames the problem, supporting driver awareness (e.g. drunk driving prevention, seatbelt usage programs), legislation, and direction in prioritizing the allocation of research dollars.

Naturalistic data can support these macro-level analyses. For example, the analysis of crash causal factors based on the 100-Car data was summarized in a series of tree diagrams (Figure 3) that illustrate the prevalence of diverse contributing and associative crash factors for each crash type observed. The most common type of conflict type resulting in a crash was Single Vehicle (~35%); Lead-Vehicle conflict was the most common (~50%) for near-crashes.

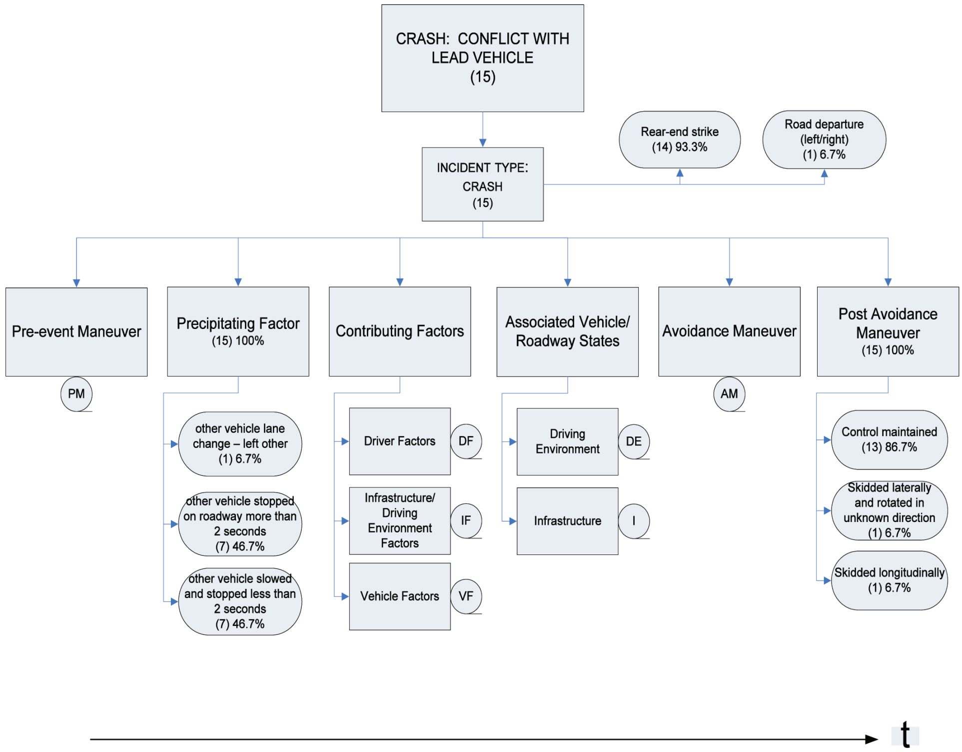

Figure 3: A sample tree diagram from the 100-Car Study

Beyond the potential of naturalistic driving data to supplement the information in crash databases, however, these data allow for the analysis of the crash causation problem at the micro level, where a particular driver and/or driving situation are analyzed. In this case, one goal may be to quantify the risk that a particular driver is exposed to during their trip, which is much more difficult than quantifying crash causal factors at a macro level. Having naturalistic observations of actual crashes and near crashes makes possible to establish links between driver behaviors and crash or near crash risk.