Title
False Alarms and Human-Machine Warning Systems

Authors
Zabyshny, Aleksandr A
Ragland, David R

Publication Date
2003-08-01

False Alarms and Human-Machine Warning Systems

Aleksandr A. Zabyshny
University of California Traffic Safety Center
140 Warren Hall
Institute of Transportation Studies and School of Public Health
University of California at Berkeley
Berkeley, CA 94720-7360
Phone: (510) 642-0655; Fax: (510) 643-9922
aaz@uclink.berkeley.edu

David R. Ragland
University of California Traffic Safety Center
140 Warren Hall
Institute of Transportation Studies and School of Public Health
University of California at Berkeley
Berkeley, CA 94720-7360
Phone: (510) 642-0655; Fax: (510) 643-9922
davidr@uclink.berkeley.edu

ABSTRACT

This work illustrates that false alarms are likely to have significant frequencies as well as detrimental influence on the effectiveness of human-machine warning systems. Several factors are responsible for false alarm materialization, including the need to predict uncertain conditions in the future, variability of human perception, and low a priori probabilities of traffic collisions. The effect of false alarms on human trust in warning systems and on credibility of warnings could be considerable even for low false alarm rates. One way to decrease false alarm rates would be to focus on predicting possible conflicts, which are much more probable than actual collisions (thus increasing a priori probabilities). One of the suggested future research directions is to assess directly what different individuals will view as being a false alarm. In the majority of existing studies, the researchers defined this outcome. It is possible that people will change their behavior based on their perception of false alarms rather than on a researcher's definition of them.

INTRODUCTION AND BACKGROUND

Motor vehicle crashes were the cause of nearly forty-two thousand deaths in year 2000 *(1)*. The total economic cost of all crashes exceeded two hundred and thirty billion dollars, which is more than two percent of the U.S. Gross Domestic Product *(2)*. The sheer impact of traffic collisions on society warrants research into safety improvements that would decrease the frequency of traffic crashes or at least reduce their severity.

Intelligent transportation systems (ITS) appear to have a sound potential to apply emerging technologies to increase safety. Throughout the past several years, numerous systems have been proposed and more research is in progress *(3)*, *(4)*, and *(5)*. The researchers often stress the importance of looking from the human factors perspective when evaluating warning systems and the expected effectiveness of such systems in reducing collisions *(6)*, *(7)*, *(8)*, and *(9)*.

The focus of the following discussion is the introduction of systems that detect potentially dangerous conditions and then warn vehicle operators of possible threats. It is important to stress that these systems are not designed to take control of the driving task as discussed by Chovan *(10)*, *(11)*, and *(12)*. Even when such systems are operational, drivers still are the ones who will be ultimately responsible for making driving decisions. A schematic representation of information flow in such human-machine systems is shown in Figure 1.

If a warning system is not deployed, vehicle operators receive information only through information link 0-2. A warning system introduces a new 0-1-2 information link. The need for this new link may be justified if vehicle operators are unable to obtain sufficient and useful information necessary to make safer driving decisions through link 0-2 alone. A warning system must have the capacity to detect the needed information in such situations and to communicate it reliably to vehicle operators.

The following example illustrates two situations in which warning systems can be helpful to vehicle operators by providing useful information. A driver is attempting to perform a left turning maneuver while vehicles that may be approaching from the opposite direction are not required to stop (e.g. permissive traffic signal indication or a two-way stop sign controlled intersection). A diagram of the scenario is provided in Figure 2. In the first situation, the driver that is turning left is unable to see approaching vehicles (due to queued up vehicles in the opposing left-turn lane, intersection geometry, etc.). The second scenario applies to a driver that is able to see the approaching vehicles but fails to evaluate the gap appropriately. The example illustrates a common traffic situation. Crashes of vehicles identified by these pre-crash maneuvers constitute more than six percent of all traffic collisions in the United States. The nomenclature and in-depth description of similar traffic crashes may be found in *(13)*.

For the benefit of the subsequent analysis, it is important to define the scope of the discussion. A human-machine system that is designed to utilize automation in order to acquire and convey additional useful information to human operators may be viewed as consisting of two subsystems: detection subsystem and information management subsystem, somewhat comparable to the structure described by Breznitz *(14)*. The detection subsystem uses sensors to recognize certain parameters (e.g. position, speed, and acceleration of vehicles approaching an intersection) that usually identify with dangerous situations to be avoided. After detection, the subsystem issues warnings alerting the drivers to these possible dangers, using preset rules (e.g., a warning system issues an alert in a manner that is timely and reliably understood by a vehicle operator). These warnings constitute the additional information that either may or may not be used by vehicle operators.

A driver may be depicted as an information management subsystem that is constantly receiving information from the environment and from the detection subsystem. Based on all information, this person has to make decisions. In the examples above, a vehicle operator must decide whether to initiate the left-turn maneuver or wait until the approaching vehicles have passed.

Therefore, even if the effectiveness of the detection subsystem can be estimated based on the technical parameters of the equipment, the detection algorithms, and on the environmental characteristics, the evaluation of effectiveness of the human-machine system (warning subsystem and information management subsystem together) will require additional insights into the effect of automation on human operators. More specifically, the influence of automation errors on credibility and trust that drivers attach to automated warning systems is of great concern.

The motivation of this paper is to address false alarms and their influence on human reaction to the information that is generated by detection subsystems. This reaction constitutes the false alarm effect (FAE) or, as it has been addressed in some sources, "cry wolf" phenomenon.

SOURCES OF FALSE ALARMS

Numerous factors affect the credibility of human-machine warning systems. These factors include relevance and timeliness of warnings. To issue a warning, a detection subsystem must evaluate the likelihood of occurrence of a certain dangerous event into the future. The longer the time into the future, the less accurate the prediction could be, and thus, the less relevant the warning will be to a receiver. On the other hand, it follows that the closer in time the dangerous event, the higher the predictive power of the warnings issued by the subsystem. However, at some point the detection subsystem, although predicting the danger with high probability, will cease to be useful to drivers. As timing of warnings moves closer to the beginning of the dangerous event, vehicle operators will not have enough time to reliably use the additional information to take protective measures. Hence, the condition of warning timeliness is not satisfied. As a reference, it is valuable to note that driver perception-reaction times vary from fractions of a second to more than three seconds. These times depend on whether the warning is expected or not, the complexity of decision a driver needs to make, and on the driver characteristics, such as age *(15)*.

Decreased relevancy and timeliness of alerts generated by detection subsystem could decrease the credibility of the subsystem in the eyes of the users. In other words, users would treat the information coming through link 0-1-2 in Figure 1 as nearly worthless.

A related consideration is a trade-off between misses and false alarms. Often, designers set the criteria for warning activations to minimize misses. Such practice typically increases the probability of false alarms. Signal detection theory (SDT) tools are available to estimate the optimum detection threshold (criterion), which minimizes the total cost of possible errors *(16)* and *(17)*, assuming that people that receive warnings would follow the system advice in all instances. Misses generally have very large costs so the criterion is usually set to a value that results in the rate of misses being far lower that the rate of false alarms.

In evaluating the effectiveness of human-machine warning systems, it is important to address the inherent variability of human perception of the same conditions in the environment. One of the conditions relevant to this discussion is the notion of gap acceptance, which may be defined in terms of arrival time *(18)*. The same gap may be too liberal for one person while being too conservative for another. This manifests itself in people using their vehicles' horns to announce to the world that they would have accepted gaps that people in front of them just rejected. The safety buffers built into timing of algorithms of detection subsystems to accommodate the majority of driving population may contribute to similar effects.

To illustrate the above point, consider a detection subsystem that is designed so that the driver preparing to turn left at an intersection receives a warning if vehicles approaching the intersection from the opposite direction are within a predetermined number of seconds. This time may be objectively set so that the majority of drivers would be accommodated. An ideal detection subsystem would acquire the danger and issue a warning with perfect reliability. However, from the point of view of the driver who is a receiver of the warning, the scenario could be far from ideal. The driver would receive a warning, take protective measures (slowing down or stopping) in anticipation of the arrival of the vehicles and they indeed arrive but too late: the driver would have accepted a much smaller gap. This driver may consider that the system has generated a false positive and that this erroneous warning just cost him or her an opportunity to turn.

The false positive in this scenario is a subjective false positive, as it is predominantly dependent on the drivers' internal characteristics. Individual differences among drivers, such as personality traits as well as drivers' perception of initial credibility of the system, determine the susceptibility to the effect of false alarms and are likely to have a marked effect of the frequency of subjective false alarms *(14)*.

Another complication is that even for near perfect detection subsystems (i.e. very low probability of misses and false alarms), the performance of human-machine systems may be only marginal. Low a priori probabilities (base rates) of traffic collisions are to blame for that. Low base rate often results in low posterior probabilities of warnings. Posterior probabilities of warnings characterize likelihood that a particular warning represents a valid potential crash or conflict *(19)* and *(20)* or, in other words, the posterior probability is the probability of a dangerous condition given a warning was issued.

This fact could have a profound effect on the effectiveness of human-machine warning systems. Perhaps, the effect of low a priori probability of traffic collisions on the false alarm rate could be better illustrated in the following example: Assume that there is a detection subsystem that could be characterized by a high rate of true positives, say 0.999 and a low rate of false positives, say 0.001. If the a priori probability of a signal is 0.0002 (as a reference, Farber and Paley, *(21)*, estimated that the base rate for highway rear-end collisions was 0.000173). The posterior likelihood of a true alarm for such detection subsystem is about 0.17. Given such posterior probability, a human driver would encounter about one true alarm out of every six warnings.

HUMAN REACTION TO FALSE ALARMS

It appears that false alarms are unavoidable artifacts of human-machine warning systems. The discussion then turns to a very important question: What is the impact of a high number of false alarms on driver response to the warnings?

The adverse reaction of humans to false alarms is likely to have deep psycho-physiological roots, as indicated by Breznitz *(14)*. Breznitz used changes in heart rates and skin conductance, among other measures, to evaluate the reaction of human organism to false alarm occurrences. His extensive experiments showed that human responses to false alarms include reductions in probability of engaging in protective behavior, reductions in protective behavior intensity, and increases in latency between the warning and the beginning of taking protective measures.

A large body of literature exists that addresses the effects of false alarms on human operators. The large majority of works concentrate on systems that are not directly relevant to automobile applications. Nevertheless, these works offer insights into far-reaching effects of automated warning systems' errors on human operators *(22)*, *(20)*, and *(23)*.

Several studies that illustrate false alarm effect on automotive collision avoidance warnings are described below. Chugh and Caird, *(24)*, explored the effects of reliability and failures of in-vehicle train warnings on driver perception-response time and trust in the warning system. Using video simulations of approaches to railroad crossings, thirty-six younger drivers were found to adjust their behavior in response to the false alarms. System trust was decreased by the decreases in system reliability and the perception-response times were increased after experiencing false alarms. A small number of people did not respond to warnings at all after being exposed to false alarms. The warnings issued by the system could have been perceived as carrying less useful information to the drivers. The reliability of the warning system was controlled by the researchers and assumed two discrete values of 83 and 50 percent.

A study by Dingus et al., *(25)*, evaluated the effect of false alarms on drivers' following behavior and on driver perception of system reliability. Forty people were separated into two groups based on their age and presented with various false alarm rates. The participants were exposed to false alarm rates of 0 percent to 30 percent, 31 percent to 60 percent, and greater than 60 percent. While the experiment was conducted using naturalistic data collection, false alarms were triggered by the experimenter, possibly influencing the results. As the false alarm rate exceeded 60 percent, younger drivers showed changes in behavior, possibly indicating a decrease in trust in the warning system and "decided to maintain whatever following distance they felt was appropriate regardless of the system warnings." It is worth noting that while exposed to lower false alarm frequencies, the younger drivers reacted in a way consistent with drivers receiving no false alarms. There was no false alarm effect (positive or negative) observed for older drivers.

Lehto et al., *(26)*, performed an experiment to evaluate the effectiveness of a collision warning system using an interactive simulation of a rural two-lane driving environment. Fifteen drivers needed to make decisions whether to pass a slower moving vehicle ahead of them. It was found that increased rate of false alarms had a detrimental effect on the risk taking behavior (increased risk taking). The authors alluded to the hypothesis that over-warning can be worse than under-warning and that the effects of over-warning were worse than expected. Authors indicate that additional research is necessary to confirm their hypothesis.

Ben-Yaacov et al., *(27)*, used naturalistic data collection to assess the effect of imperfect warning system on driver headway maintenance and on driver behavioral response to warning system errors. The experiment showed that for the range of reliabilities from 60 to 95 percent there were no significant changes in driver behavior. The authors found it reassuring that the warning system does not have to be perfect to be useful.

While some studies indicate that the FAE is insignificant or nonexistent, one must keep in mind that the false alarm rates used in these studies are relatively low. As it was described in the previous section, the rate of false alarm occurrences may be considerably higher than the rates applied in the experiments. Even considering the low frequencies of false alarms used in the studies, most experiments show that these errors may significantly affect the effectiveness of the human-machine warning systems. Thus, the perceived reliability of the automation often influences the value people place on the system warnings *(28)*. This fact warrants consideration of possible measures that could reduce false alarm rates or alleviate the impact of such errors.

COUNTERMEASURES

Besides increasing the sensitivity of sensors and detection algorithms, a way to reduce the frequency of false alarms could lie in redefining the dangerous event. Instead of striving to predict actual collisions, which are very rare events, the detection subsystem could focus on only predicting possible conflicts (e.g., the system would issue a warning if vehicles approaching the intersection from the opposite direction were within a certain time from entering the intersection). These events are not nearly as rare as collisions and, thus, the posterior probabilities of warnings are much higher limiting the rate of false positives.

After all, not all false alarms are created equal. Under some circumstances, false alarms may be beneficial for human-machine performance. Farber and Paley contemplated that the algorithm should issue warnings in collision-possible circumstances, although a driver would likely be able to complete an avoidance maneuver successfully. Although in principle a false alarm, this type of warning could be interpreted as an aid in allowing improved response to an alarm in a collision-likely situation. On the other hand, the information that is gained by drivers through alarms that warn only about approaching vehicles rather than impending collisions may be less appreciated by the drivers, as carrying less value.

Nevertheless, it appears that false alarms are to reside in the world of human-machine warning systems. While there is no complete defense against the effect of false alarms, the influence of them may be reduced. Breznitz suggests that pre-defined rules requiring taking protective measures after a warning would reduce the effect of false alarms on humans. Such rules would also lessen the pressure on the drivers to make complex decisions as well as would allow them to engage in safer behavior even if it is unnecessary on some occasions. The clear way to implement this countermeasure is to require vehicle operators to take or not to take certain predefined actions after each warning (e.g. wait until the warning is cancelled before initiating the left-turn maneuver). This countermeasure, however, may limit the applicability of some warning systems, such as in the case in which conventional crash countermeasures are available (e.g. installation of a protected left-turn phase, instead of deploying left turn warning devices).

CONCLUSIONS

The number of traffic collisions and their cost to society warrant the development of ITS countermeasures to reduce the rate of these crashes. It is important not to underestimate the effect of human interaction with automation in assessing the effectiveness of human-machine warning systems. False positive rate could be significant and false alarms are likely to adversely affect the performance of such systems.

The design of physical components and systems is considered to be a dominant engineering task. However, the design decisions made during the design stage “can forever plague the user” and cause lower than expected human-machine warning system performance if the human factors are not considered. On the other hand, the proper use of relevant human factors in the design could facilitate the use and bring enhanced system performance.

Different people may perceive a potentially dangerous situation differently. Drivers could exhibit different levels of trust in the system. At two possible extremes, drivers could over-rely on the warning system or have no trust in it *(29)*. Additional research is necessary to gain insights into the extent of influence of such conditions on the effectiveness of human-machine systems.

One of the important considerations in setting detection thresholds is the fact that the perception of what is considered to be a false alarm may vary from driver to driver and even for any individual driver in different situations. The level of variability among drivers could be a vital factor in assessing the effectiveness of the human-machine systems. In the studies reviewed, the researchers defined false alarms and false alarm rates themselves. Perhaps it would be beneficial to study directly what different individuals will view as being a false alarm. It is possible that people will change their behavior based on their perception of these outcomes rather than on a researcher's definition of them.

REFERENCES

1. NHTSA, *Traffic Safety Facts 2000: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*, DOT HS 809 337. 2001, National Highway Traffic Safety Administration: Washington, D.C.
2. Blincoe, L*., et al.*, *The Economic Impact of Motor Vehicle Crashes, 2000*, DOT HS 809 446. 2002, National Highway Traffic Safety Administration: Washington, D.C.
3. Wilson, T., *IVHS Countermeasures for Rear-End Collisions, Task 1 Interim Report: Volume VI: Human Factors Studies*, DOT HS 808 565. 1994, Frontier Engineering: Advanced Program Division for National Highway Traffic Safety Administration: Scottsdale, AZ.
4. USDOT, *Infrastructure Systems: Intersection Collision Avoidance -- Overview of Concepts*. 2001, U.S. Department of Transportation.
5. USDOT, *Intelligent Vehicle Initiative: Business Plan*. 2000, U.S. Department of Transportation: Washington, D.C.
6. Batelle, *Full-Coverage Collision Warning: Human Factors Research Needs*, FHWA-RD-98-185. 1998, Battelle Human Factors Transportation Center for Federal Highway Administration: Seattle, Washington.
7. Papastavrou, J.D. and M.R. Lehto, *Improving the effectiveness of warnings by increasing the appropriateness of their information content: some hypotheses about human compliance.* Safety Science, 1996. **21**: p. 175-189.
8. Lerner, N.D*., et al.*, *Inappropriate Alarm Rates and Driver Annoyance*, DOT HS 808 533. 1996, National Highway Traffic Safety Administration: Washington, D.C.
9. Parasuraman, R. and M. Mouloua, *Automation and Human Performance: Theory and Applications*. 1996, Mahwah, New Jersey: Lawrence Erlbaum Associates.
10. Chovan, J.D*., et al.*, *Examination of Intersection, Left Turn Across Path Crashes and Potential IVHS Countermeasures*, DOT HS 808 164. 1994, John A. Volpe National Transportation Systems Center for National Highway Traffic Safety Administration: Washington, D.C.
11. Chovan, J.D*., et al.*, *Examination of Unsignalized Intersection, Straight Crossing Path Crashes and Potential IVHS Countermeasures*, DOT HS 808 152. 1994, National Highway Traffic Safety Administration: Washington, D.C.
12. Chovan, J.D*., et al.*, *Examination of Signalized Intersection, Straight Crossing Path Crashes and Potential IVHS Countermeasures*, DOT HS 808 143. 1994, National Highway Traffic Safety Administration: Washington, D.C.
13. Najm, W.G., J.D. Smith, and D.L. Smith, *Analysis of Crossing Path Crashes*, DOT HS 809 423. 2001, John A. Volpe National Transportation Systems Center for National Highway Traffic Safety Administration: Washington, D.C.
14. Breznitz, S., *Cry Wolf: The Psychology of False Alarms*. 1984, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
15. AASHTO, *A Policy on Geometric Design of Highways and Streets*. 4th ed. 2001, Washington, D.C.: American Association of State Highway and Transportation Officials. 905.
16. Helstrom, C.W., *Statistical Theory of Signal Detection*. International Series of Monographs on Electronics and Instrumentation, ed. D.W. Fry and W.A. Higinbotham. Vol. 9. 1960, London: Pergamon Press. 364.

17. Parasuraman, R., A.J. Masalonis, and P.A. Hancock, *Fuzzy Signal Detection Theory: Basic Postulates and Formulas for Analyzing Human and Machine Performance.* Human Factors, 2000. **42**(4): p. 636-659.
18. Caird, J.K. and P.A. Hancock, *The Perception of Arrival Time for Different Oncoming Vehicles at an Intersection.* Ecological Psychology, 1994. **6**(2): p. 83-109.
19. Parasuraman, R., P.A. Hancock, and A. Olofinboba, *Alarm effectiveness in driver-centred collision-warning systems.* Ergonomics, 1997. **40**(3): p. 390-399.
20. Meyer, J. and Y. Bitan, *Why Better Operators Receive Worse Warnings.* Human Factors, 2002. **44**(3): p. 343-353.
21. Farber, E. and M. Paley. *Using freeway traffic data to estimate the effectiveness of rear end collision countermeasures*. in *Third Annual IVHS America Meeting*. 1993. Washington, D.C.
22. McCormick, E.J. and M.S. Sanders, *Human Factors in Engineering and Design*. 5 ed. 1982, New York, New York: McGraw-Hill. 615.
23. Cafarelli, D.A., *Effect of False Alarm Rate on Pilot Use and Trust of Automation Under Conditions of Simulated High Risk*, in *Department of Aeronautical and Astronautical Engineering*. 1998, Massachusetts Institute of Technology: Cambridge. p. 49.
24. Chugh, J.S. and J.K. Caird. *In-Vehicle Train Warnings (ITW): The Effect of Reliability and Failure Type on Driver Perception Response Time and Trust*. in *Human Factors and Ergonomics Society 43rd Annual Meeting*. 1999. Houston, TX: Human Factors and Ergonomics Society.
25. Dingus, T.A*., et al.*, *Human Factors Field Evaluation of Automotive Headway Maintenance/Collision Warning Devices.* Human Factors, 1997. **39**(2): p. 216-229.
26. Lehto, M.R*., et al.*, *An experimental comparison of conservative versus optimal collision avoidance warning system thresholds.* Safety Science, 2000. **36**(3): p. 185-209.
27. Ben-Yaacov, A., M. Maltz, and D. Shinar, *Effects of an In-Vehicle Collision Avoidance Warning System on Short- and Long-Term Driving Performance.* Human Factors, 2002. **44**(2): p. 335-342.
28. Parasuraman, R. and V. Riley, *Humans and Automation: Use, Misuse, Disuse, Abuse.* Human Factors, 1997. **39**(2): p. 230-253.
29. Wickens, C.D., *Engineering Psychology and Human Performance*. 2 ed. 1992, New York, New York: HarperCollins Publishers. 560.

LIST OF TABLES AND FIGURES

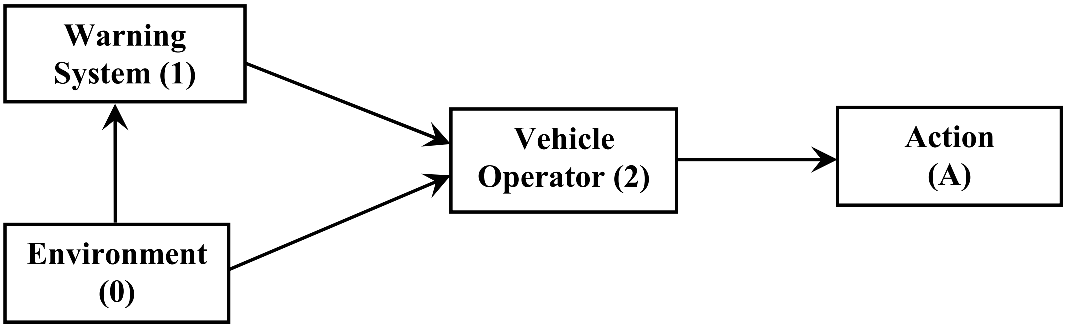

FIGURE 1 Simplified schematic showing information dynamics.

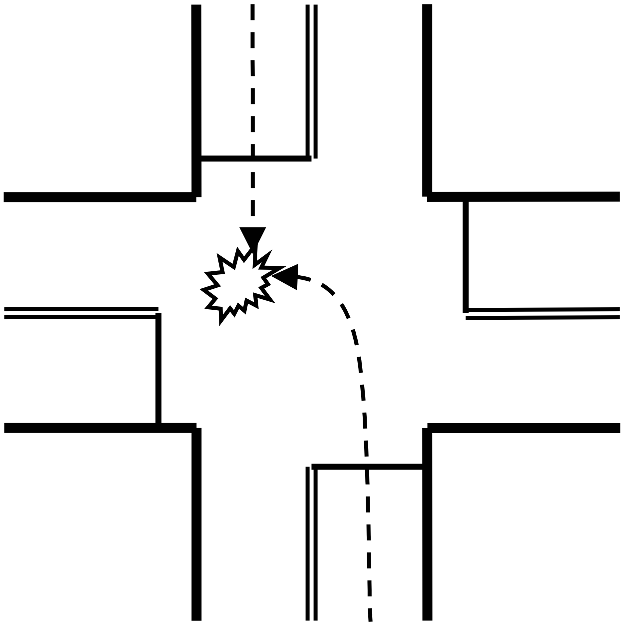

FIGURE 2 Schematic Representation of a Left Turn Across Path/Opposite Direction Conflict Scenario.