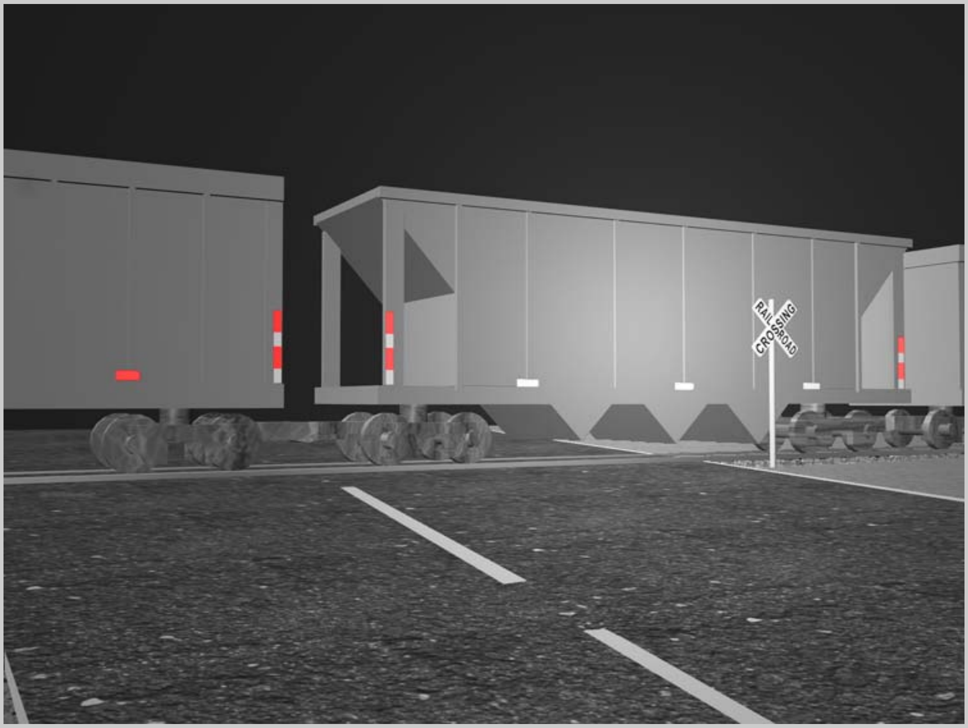
U.S. Department
of Transportation
**Federal Railroad
Administration**

Effects of Active Warning Reliability on Motorist Compliance at Highway-Railroad Grade Crossings

**Office of Research
and Development
Washington DC 20590**



Human Factors in Railroad Operations

always expect a train

DOT/FRA/ORD-09-06

Final Report

This document is available to the

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (*LEAVE BLANK*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November, 2007	September 2003- June 2006

4. TITLE AND SUBTITLE Effects of Active Warning Reliability on Motorist Compliance at Highway-Railroad Grade Crossings	5. FUNDING NUMBERS RR04/DB105
6. AUTHOR(S) Monica Gil, Jordan Multer, and Michelle Yeh	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Department of Transportation	8. PERFORMING ORGANIZATION DOT-VNTSC-FRA-09-04
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Department of Transportation	10. SPONSORING/MONITORING AGENCY REPORT NUMBER DOT/FRA/ORD-09-06
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT This document is available to the public through the National Technical Information Service, Springfield, VA 22161	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Federal Railroad Administration is interested in understanding the effect of warning reliability on motorist compliance to warnings at active grade crossings. The John A. Volpe National Transportation Systems Center conducted two studies to examine motorist behavior as a consequence of warning signal failure. Experiment one measured motorist behavior in response to false alarms (i.e., the presentation of a warning when no train was approaching). Experiment two examined how motorist responses to grade crossing warning signals were influenced by false alarms and missed signals (i.e., the failure of the warning system to signal an approaching train).

The results of the experiments support the hypothesis that warning systems reliability exerts a predictable effect on motorists. As motorist' perceive the warning system to be less reliable, they are more likely to violate the warning signal, perceiving little risk to their safety. Although it is not possible to define empirically the precise warning reliability required to achieve a desired level of compliance, measures for improving motorist behavior from an engineering and cognitive science perspective are considered and areas for future research are discussed.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Warning reliability, highway-rail grade crossing, traffic control devices, grade crossing safety, active crossings, motorist compliance	89
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unclassified

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in)	= 2.5 centimeters (cm)
1 foot (ft)	= 30 centimeters (cm)
1 yard (yd)	= 0.9 meter (m)
1 mile (mi)	= 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2)	= 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2)	= 0.09 square meter (m^2)
1 square yard (sq yd, yd^2)	= 0.8 square meter (m^2)
1 square mile (sq mi, mi^2)	= 2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he)	= 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz)	= 28 grams (gm)
1 pound (lb)	= 0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb)	= 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp)	= 5 milliliters (ml)
1 tablespoon (tbsp)	= 15 milliliters (ml)
1 fluid ounce (fl oz)	= 30 milliliters (ml)
1 cup (c)	= 0.24 liter (l)
1 pint (pt)	= 0.47 liter (l)
1 quart (qt)	= 0.96 liter (l)
1 gallon (gal)	= 3.8 liters (l)
1 cubic foot (cu ft, ft^3)	= 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3)	= 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] °F	= y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm)	= 0.04 inch (in)
1 centimeter (cm)	= 0.4 inch (in)
1 meter (m)	= 3.3 feet (ft)
1 meter (m)	= 1.1 yards (yd)
1 kilometer (km)	= 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2)	= 0.16 square inch (sq in, in^2)
1 square meter (m^2)	= 1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2)	= 0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2)	= 1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm)	= 0.036 ounce (oz)
1 kilogram (kg)	= 2.2 pounds (lb)
1 tonne (t)	= 1,000 kilograms (kg)
	= 1.1 short tons

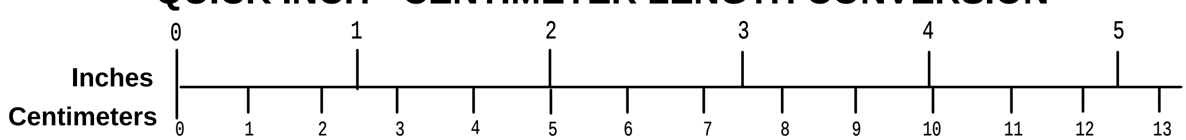
VOLUME (APPROXIMATE)

1 milliliter (ml)	= 0.03 fluid ounce (fl oz)
1 liter (l)	= 2.1 pints (pt)
1 liter (l)	= 1.06 quarts (qt)
1 liter (l)	= 0.26 gallon (gal)
1 cubic meter (m^3)	= 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3)	= 1.3 cubic yards (cu yd, yd^3)

TEMPERATURE (EXACT)

[(9/5) y + 32] °C	= x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSION

°F	-40°	-22°	-4°	14°	32°	50°	68°	86°	104°	122°	140°	158°	176°	194°	212°
°C	-40°	-30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°

For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50
SD Catalog No. C13 10286

Updated 6/17/98

Acknowledgements

This research was conducted with funding from the Federal Railroad Administration's (FRA) Office of Research and Development. The authors wish to thank Dr. Thomas Raslear of the FRA Office of Research and Development for his direction and helpful guidance in the development of the experiment and report. Special thanks also go to Prism Technologies for their expertise in developing the simulation and to the many people who participated in the study.

The views expressed herein are those of the authors and do not necessarily reflect the views of the John A. Volpe National Transportation Systems Center, the Research and Innovative Technology Administration, or the United States Department of Transportation.

Table of Contents

Illustrations

Tables

Executive Summary

The Federal Railroad Administration (FRA) is interested in understanding the effect of warning reliability on motorist compliance to signals at active grade crossings. When the active warning device is activated, the motorist is generally required to stop at the crossing, look for a train, and if one is spotted, wait for it to pass. However, several factors contribute to motorist noncompliance, including low signal reliability, the motorist's low expectancy for a train at a particular crossing derived from the motorist's familiarity with the crossing, and inconvenience due to long waiting times. Of interest is the issue of warning reliability and the degree to which warning signal failure affects motorist compliance at grade crossings.

The FRA sponsored the John A. Volpe National Transportation Systems Center to conduct two studies to examine motorist behavior because of unreliable warning signals. Experiment 1 examined motorist behavior in response to warning false alarms (i.e., the presentation of a warning when no train was approaching). Experiment 2 examined how motorist responses to grade crossing warning signals were influenced by false alarms and missed signals (i.e., the failure of the warning system to signal an approaching train). The experimental methods, results, and their implications are discussed in the following sections.

Experiment 1: Static Task

In Experiment 1, we assessed whether participants were sensitive to reduced warning reliability. Participants viewed a series of static images of actively protected highway grade crossings and made a decision regarding whether to stop or proceed. The images showed an activated grade crossing warning device with the gate in the lowered position and the red light on. We manipulated the reliability of the warning system by varying its positive predictive value (PPV), the probability that the warning truly indicated a dangerous condition, at eight levels: .23, .30, .40, .60, .70, .77, .87, and .97. Participants' performance was measured as a function of the proportion of valid stops and proportion of false stops. A *valid stop* was defined as the case when a warning signal is presented (and reliable) and the motorist stops at the crossing. A *false stop* was defined as the case when a warning signal is presented but it is unreliable and motorist stops at the crossing unnecessarily. A feedback screen provided participants with information about whether their response was correct or incorrect after they made their decision.

We analyzed the data using signal detection theory to examine participants' sensitivity and response bias as a function of warning reliability. We also conducted a proportional analysis to determine if compliance was influenced by warning reliability systematically. The results of both analyses showed that participants' likelihood to comply dropped as warning reliability decreased (i.e., as the PPV rate dropped). Participants were not sensitive to changes in warning reliability, unless the PPV rate was high (e.g., a drop from .97 to .87). An examination of participants' shift in response bias also indicated that participants were more likely to exhibit risky behavior by proceeding rather than stopping when confronted with an ambiguous grade crossing situation and a low PPV rate.

Thus, participants tended to match their responses to the PPV rate, such that they were more likely to comply when warning reliability was high. It is important to note that because participants exhibited compliant behavior with reliable warning systems, as reliability increased, so did participants' false stopping responses. From a traffic law standpoint, this behavior is desirable, but from a human factors perspective, these false stops contribute to motorist frustration at grade crossings and facilitate mistrust in the warning system. Consequently, in the future, motorists may distrust an accurate active warning signal and engage in gate violation behavior, leading to an accident or near miss.

One limitation of Experiment 1 is that it examined motorist behavior using a static environment and focused only on the impact of false alarms on motorist behavior. We conducted Experiment 2 to examine the effect of system *un*reliability in a more realistic dynamic driving environment and to evaluate the impact of false alarms *and* missed signals on motorists' decisions at grade crossings.

Experiment 2: Driving Task

In Experiment 2, we used a signal detection paradigm to evaluate motorist compliance with actively protected grade crossings in a simulated driving environment. Participants performed two tasks: a *priming task* to set their expectations about warning system reliability and a *driving task* in a simulator. In the priming task, participants viewed a series of static images of gated highway-railroad grade crossings and determined whether to stop or proceed. The images were similar to those shown in Experiment 1 but images of inactive warning signal devices were also included with the images of active warning signals. Additionally, a train horn sound was incorporated to indicate train arrival.

In the driving task, participants drove a simulated vehicle through a course with 24 active grade crossings with partially reliable warning systems. The design of the simulated grade crossing environment was similar to an actual grade crossing with one exception—the second gate arm was omitted from the design of the simulated crossings to simplify participants' maneuvering around lowered gate arms in the simulator when they chose to do so.

Participants completed the priming task before the driving task, and they completed both tasks for one warning reliability level before moving on to the next one. Warning reliability, as measured by the PPV rate, was manipulated by varying the rate of false alarms and misses. The PPV rate was set at three levels: .40, .60, and .83. Participants' performance in both the priming and driving tasks were measured by their rate of compliance, their sensitivity, and response bias. In the driving task, participants were also evaluated using their collision frequency, task completion time, and train time to crossing.

In first considering the priming task, the results indicated that compliance increased as PPV rate increased when the warning system was reliable. Unlike Experiment 1, participants were sensitive to the accuracy of the warning system as reliability improved. In other words, participants became better able to distinguish reliable from unreliable warnings. One change to the methodology used in Experiment 1 for the priming task in Experiment 2, which could account for this difference in sensitivity, was the addition of the sound of a train horn to indicate imminent train arrival. The train horn provided an auditory cue in conjunction with the visual cue of the lowered gate, but unlike the visual warning, the auditory warning was perfectly reliable.

Examination of participants' response bias in the priming task showed that they were generally conservative in their responses—and likely to stop at the crossing—even when the reliability of the warning system was low. However, once participants were in the simulator, a different pattern of behavior emerged. Participants were inclined to proceed (a liberal criterion) regardless of the reliability level. Examination of participants' sensitivity level showed that they were sensitive to the PPV rate in the driving task when warning reliability was high, but as the PPV rate dropped, participants' became less able to distinguish between reliable versus unreliable warnings.

Of the descriptive driving performance measures, data for the frequency of gate violations showed the most compelling evidence of the costs of unreliable warnings. As the PPV rate decreased, the frequency of gate violations increased. Interestingly, a comparison of driving task completion time indicated that violating the gates did not significantly reduce the time required to reach the destination, despite the fact that a few participants drove through the course without complying with any of the warning signals.

The results of Experiment 2 support the hypothesis that warning system unreliability can have a detrimental but predictable effect on motorists. As motorists' perceive the warning system to be less credible, they will be more likely to violate the warning signal, perceiving little risk to their safety since the warning system has failed before.

<u>General Discussion and Future Directions</u>

The results of Experiments 1 and 2 suggest that improving motorists' perception of signal reliability may improve compliance. Motorists were sensitive to the reliability of the active warning devices, particularly in the driving simulator, and they were more likely to comply when they perceived the warning to be reliable. Unfortunately, it is not possible to empirically define the precise warning reliability required to achieve a desired level of compliance. From an engineering perspective, the false alarm rate can be reduced through improvements in track circuitry and train detection equipment, incorporating good maintenance practices, and identifying and correcting signal malfunctions in a timely manner. From a cognitive science perspective, additional research is needed to investigate factors that motorists use to judge warning system credibility.

Based on the results of the current experiments, we recommend the following areas for research:

- Examine the value provided by different external cues regarding a train's arrival at the crossing (e.g., the sounding of a train horn),
- Develop a decision-making model of compliance based on the expected value of information when a warning is presented and the expected value of information when no warning is presented,
- Examine the interaction between the motorist and warning signal using a model of distributed team signal detection,
- Understand motorists' cost-benefit structures that determine their response at a crossing, and

- Investigate how motorists' expectancies regarding the likelihood of a train at a crossing factors in compliance.

1. Introduction

1.1 Problem Statement

The current research addresses motorist compliance at active crossings protected by flashing lights and two quadrant gates. At these crossings, the motorist receives information from warning devices positioned at the crossing about whether a train is approaching, and under ideal conditions, does not need to determine whether a train is approaching. When the active warning device is activated, it is the motorist's responsibility to stop at the crossing and wait for the train to pass. However, a motorist's expectancies of the credibility of the warning device and the length of the warning time, developed from past experiences, factor in the decision to comply. Long waiting times and unreliable warning signals are believed to exert detrimental effects on motorist compliance at active grade crossings. Thus, the Federal Railroad Administration (FRA) was interested in empirically examining the effect of warning reliability on motorist compliance.

This report describes two studies that investigated motorist behavior with flashing lights and gates at highway-railroad grade crossings and examined the effects of unreliable warning signals. We use the term "warning reliability" to refer to the degree to which a warning accurately predicts the presence of a train at the grade crossing. We begin by reviewing literature examining motorist behavior at active grade crossings. Next, we address the design of warning systems and the effects of warning *un*reliability on operator behavior. Since only a few studies specifically examined motorist compliance to warnings in the grade crossing domain, the general literature on warnings is discussed, as applicable.

1.2 Motorist Behavior at Active Grade Crossings

At an active grade crossing, warning devices positioned at the crossing present the motorist with information about the likelihood of an approaching train. The train triggers the onset of flashing lights, usually with a minimum operation time of 20 seconds before the arrival of the train at the crossing. When the crossing is also protected with automatic gates, the gate arms lower to a height approximately 4 ft from the pavement to block highway traffic. Many crossings also incorporate audible bells for alerting pedestrians and cyclists to supplement the lights and gates.

When the active warning device is activated, the motorist has a responsibility to stop at the crossing and wait for the train to pass. Although motor vehicle code regulations vary from state to state, compliant behavior at active grade crossing with flashing lights and gates generally requires the motorist to stop at the onset of the flashing lights, and remain stopped, until the gate has been raised. Failure to fulfill these requirements results in noncompliance. Noncompliance is not considered intentional if a motorist fails to detect the traffic control devices or the crossing itself, does not comprehend the situation, or does not understand the required actions. However, it is considered intentional when a motorist consciously and deliberately ignores activated warning signals (flashing lights) and drives around the lowered gates. This type of action has been described as a "gate violation, or a "traffic violation" (Parker, Reason, Manstead and Stradling, 1995), or an "intentional unsafe act" (Caird, Creaser, Edwards, and Dewar, 2002).

In two comprehensive reviews of motorist behavior at grade crossings, Lerner et al. (1990) and Yeh and Multer (in preparation) identified several factors that could lead to noncompliance, such as low expectancy of a train's arrival derived from motorists' familiarity with the crossing, the inconvenience in terms of delays resulting from compliance, long waiting times, signal unreliability, social pressure to "beat the train" to the crossing, and lack of enforcement. Two other studies have examined motorist behavior at active grade crossings in particular to understand motorists' approach behavior and to determine why violations occur.

In one study, Meeker, Fox, and Weber (1997) compared motorist behavior at a flashing light crossing before and after it was upgraded with two-quadrant gates. The data collected consisted of the time between the onset of the warning signal and the arrival of the vehicle at the crossing, if a motorist slowed or stopped when approaching the crossing, the time it took for the motorist to clear the crossing, and the time of the train's arrival at the crossing. The results showed that fewer motorists violated the crossing after the installation of the gates; 67 percent of motorists crossed the tracks in front of an approaching train when it was protected with flashing lights, but only 38 percent of motorists violated the crossing after the installation of the gates. However, motorists who violated the crossing after the installation of the gates stopped or slowed significantly less than those who violated the crossing when it was protected with flashing lights only. Fifty-two percent of motorists who violated the gated crossing did not stop or slow down on their approach, compared to only 13 percent of motorists who violated the crossing when it was protected with flashing lights only. The finding suggests that motorists intent on violating a gated crossing determine that the "safest" way to do so is without slowing or stopping.

Abraham, Datta, and Datta (1998) observed and classified violations at active grade crossings to identify contributing factors. In their study, they observed motorists at 37 grade crossings, which were actively protected with either flashing lights or flashing lights and gates. Observers recorded violations and classified them according to one of five risk levels; *routine*, *risky*, *more risky*, *severe*, or *critical*. Definitions for each of the risk levels for each of the two protection systems and their observed rate of occurrence are presented in Table 1.

Table 1. Observed violations at active grade crossings in Abraham, et al. (1998)

Risk Level	Definition		% Observed
	Flashing Light	Flashing Lights and Gates	
Routine	Motorist violated the crossing 4 seconds or longer after the train had passed	Motorist violated the crossing after the train had passed but before the gate arms were raised and the flashing lights had stopped	27%
Risky	Motorist crossed within 4 seconds of the train's passage	Motorist crossed when the gates were still down and the lights were still flashing	33%
More Risky	Motorist crossed within 8 to 10 seconds before the arrival of the train	Motorist crossed while the gates were lowering	19%
Severe	Motorist crossed within 4 to 8 seconds before the train's arrival	Motorist maneuvered around lowered gates	19%
Critical	Motorist crossed the track with less than 4 seconds before the train's arrival	Motorists crossed the tracks when the gates were down with less than 5 seconds before the train's arrival	2%

As described in Table 1, the risk levels were defined as a function of the train's arrival time to the crossing. *Routine* and *risky* violations occurred after the train cleared the crossing. Abraham, et al. hypothesized that these violations resulted because motorists perceived a low risk in not complying with the traffic control device since the train had already passed. The other three violations occurred before the train's arrival at the crossing. Motorists who committed the *more risky* and *severe* violations tended to accelerate to clear the crossing. *Critical* violations tended to result when the train was moving slowly.

Abraham, et al. considered the observed violations in conjunction with 7 years of the crossings' crash history to determine whether the violation rates could be accounted for by the type of protection at the crossing (flashing lights only or flashing lights and gates), the number of tracks at the crossing (single or multiple), or the number of lanes on the approach (single or multiple). The results of this analysis showed that there were significantly more crashes at gated crossings with multiple tracks and multiple lanes on the approach than at flashing light crossings with single tracks and single lanes on the approach, highlighting the higher accident risk at gated crossings. Whereas one contributing risk factor may be the higher exposure level at gated crossings relative to flashing light crossings, the nature of the roadways on the approach to the crossing may also influence motorists' decision to violate a crossing. Approaches with multiple lanes provide space for motorists to maneuver around gates, and in fact, Abraham, et al. found more crashes and violations at gated crossings with multiple lanes on the approach than at those with single lanes on the approach. The results also showed that low-risk violations tended to occur at flashing light crossings with single tracks and two-lane roads on the approach. These violations were usually the result of long warning times or motorist misunderstanding of the flashing red light signal.

Thus, the results of Abraham, et al. suggest that motorist compliance may be improved by reducing the warning time. Long warning times influence motorists' perception of signal credibility, and if motorists believe a signal is not credible, then they are less likely to comply.

Motorists' perception of signal credibility may also be influenced by the warning's reliability. Generally, warnings do not have perfect reliability (i.e., 100 percent), so it is of interest to measure the degree to which warning signal failure affects motorist compliance at grade crossings. The next section reviews literature on the design of warnings and the effects of warning system failures on operator behavior.

1.3 Research on Warnings

Dynamic warnings present one of two messages, based on input from sensors (threshold algorithms): "yes" there is a signal, or "no" there is not (the inactivity of the warning is also considered to be a message). Dynamic warnings are "sensor-based signaling systems" that alert users to potential hazards and allow them to take actions that minimize risk of injury or damage (Bliss and Gilson, 1998). Dynamic warnings are also ubiquitous parts of technological systems. Typical examples are smoke alarms, hazard warnings in industrial plants, collision avoidance warnings in vehicles and aviation, alarms from monitors in intensive care, and active warnings at grade crossings.

Dynamic warnings usually comprise two components operating in tandem (Bliss and Gilson, 1998; Getty, Swets, Pickett and Gonthier, 1995; Pate-Cornell, 1986; Sorkin and Woods, 1995). The first component consists of a mechanical device that uses sensor logic (a preset decision threshold) to determine if and when to trigger a warning signal. Getty, et al. (1995) notes the challenge surrounding this component is properly setting the sensor decision threshold. A threshold that is too strict minimizes false signals, but increases the possibility that dangerous situations will go unsignaled. On the other hand, a threshold that is set too leniently will minimize these misses but the false signal rate will rise. Thus, the design of the physical components of the system must optimize the trade-offs between minimizing false warning signals and maximizing warning sensitivity (Bliss and Gilson, 1998).

The second component of a dynamic warning is the human operator, who detects, evaluates, and responds (or does not respond) to the signal generated by the warning sensor. Research consideration of the human component is far more complex than manipulating the mechanical component, as it requires an understanding of the perceptual and cognitive processes of the human operator. Often, as is the case in the experiments conducted here, the assumption is that warnings are noticed, recognized, and understood by the user, following prior instruction and/or experience. As a result, some of the issues related to the detection, recognition and comprehension of warnings will not be discussed. Instead, we focus on factors that contribute to the operator's decisionmaking process about whether or not to comply. This decision involves a consideration of the likelihood that a danger is present and a weighting of the expected costs and benefits to compliance (Lehto, 2006).

Research on warnings has often studied operator actions immediately after a warning, but it is also important to assess the consequences of warning system failure on performance. Bliss and Gilson (1998) proposed a categorization of warning signal failures with three general categories: *false signals*, *missing signals*, and *multiple signals*. The first two of these categories (false and missing signals) are relevant to the present studies. False signals result from an oversensitive sensor system, or some type of a sensor system failure, and the operator is alerted to an event that does not occur. In the grade crossing situation, a false alarm occurs when the active warning device activates even though a train does not approach. Missed signals result when the warning system fails to inform an operator about a legitimate danger. False signals and missed signals may be related to the mechanical sensor if the decision criterion is set too strictly or too leniently, respectively. While the correction of the mechanical sensors is addressed by the engineering and maintenance domains, it is also important to consider the cognitive effects of warning signal failure on human performance.

Generally, false, missing, and conflicting multiple warnings undermine confidence and trust in system accuracy, thereby reducing subsequent compliance and reliance (Breznitz, 1983; Pate-Cornell, 1986; Bliss et al., 1995). The effect of warning system failure on trust has been studied extensively by examining operators' responses to automation by explicitly measuring the operator's perception of system reliability through subjective ratings or by implicitly observing the operator's response to automation when it fails (Parasuraman and Riley, 1997).

The next two sections describe the effects of warning system false alarms (Section 1.3.1) and misses (Section 1.3.2) on operator behavior. We found only a handful of studies that directly examined the effects of unreliable warnings on motorist behavior at grade crossings, so we included research from other domains that examined the effects of warning system reliability on performance. Note that researchers often refer to warning signal terms synonymously. Standard dictionaries frequently use one term to define another, such as the term "warning" to define an "alarm" (Bliss and Gilson, 1998). Literature on sensor-based signaling systems use terms such as "warnings," "alarms," and "alerts" to indicate information provided by some sensory-based signaling state. A similar problem exists concerning the concept of warning system reliability. Here, we use the term "reliability" to refer to the quality or credibility of warning information provided to an operator. This term was adopted because it is normally employed in the research domain concerning railroad issues. Operationally, we use the term "reliability" to refer to the degree to which a warning accurately predicts the presence of a train at the grade crossing.

1.3.1 Warning System False Alarms

The detrimental effects of false alarms on performance have been documented in a variety of complex task environments such as underground mining (Mallett, Vaught, and Brnich, 1993), medical care (Bitan, Meyer, Shinar and Zmora, 2000; Kerr, 1985), and aviation (Bliss & Gilson, 1998). False alarms can lead to inappropriate responses and create a future tendency to overlook or ignore signals (i.e., the "cry-wolf phenomenon"), especially during high workload conditions (Dunn, 1995). Sorkin (1988) found that operators learned to ignore frequent false alarms and, in doing so, often ignored credible warnings. Parasuraman and Riley (1997) reported that automated alerting systems in aviation, such as the ground proximity warning system, were sometimes turned off because of the high rate of false alarms. False alarms also induce operators to respond slower to a signaled event than when alarms are completely reliable (Getty et al., 1995).

Repeated exposure to unreliable warnings at active grade crossings may diminish motorist's perception of warning reliability, so over time, motorists learn that a warning signal does not always indicate an approaching train (Lerner et al., 1990). Wilde, Hay, and Brites (1987) found that the rate of violations increased at a crossing with a high rate of false alarms relative to crossings with no false alarms. In their study, motorists at active and passive grade crossings in

Ontario were video recorded to examine the types of incidents that occurred. The reliability of warning devices at five active grade crossings (three which were protected by flashing lights and bells and two protected by gates) was collected. Whereas false alarms occurred at only one of the five active grade crossings, the number of false alarms at this crossing accounted for 50 percent of the warning signal activations. Video recordings of motorist behavior at this crossing showed a high rate of violations at the crossing when compared to the other crossings in the study. However, the observations recorded only the violation, so the causes were not known.

Chugh and Caird (1999) addressed the effect of warning information reliability more empirically in a laboratory study examining motorist compliance to an in-vehicle warning display that alerted the motorist to the presence of a grade crossing and/or train. Participants, presented with driving scenes of grade crossings in a simulator, slowed or stopped as they approached the grade crossings based on information provided by visual and auditory alerts from the in-vehicle head-up display. Participants viewed four blocks of trials. In the first two blocks, the reliability of the in-vehicle warning system was perfectly reliable (i.e., 100 percent); participants' response times to warning signals at grade crossings in these blocks was collected as a baseline. In the third block of trials, the reliability of the warning system was reduced to one of two levels: 83 percent or 50 percent. An unreliable warning consisted of one of three failures: a false alarm in which the in-vehicle warning system presented an alert but no train approached the crossing, a false alarm in which the in-vehicle warning system presented an alert but there was no crossing, or a missed signal in the system failed to warn of an approaching train (the effect of these missed signals will be discussed in Section 1.3.2). In the fourth block of trials, the reliability of the warning system was again perfectly reliable. A comparison of the baseline data collected in the first two blocks of trials and motorist reaction times to the failures in the third block of trials showed that motorists' response time to warnings increased after false alarms occurred. Response time returned to baseline levels in the fourth block of trials only when reliability was high (83 percent) in the third block of trials. Motorist's trust in the in-vehicle warning display was also measured as part of the study. The subjective results showed that trust decreased as the system became unreliable, and it decreased to a greater degree when the reliability level was 50 percent than when it was 83 percent. However, trust was quickly regained by the end of the fourth block of trials regardless of the reliability level in the third block of trials.

Whereas Chugh and Caird conducted their study in a simulated driving environment, the results of field evaluations of similar in-vehicle displays also show a high rate of false notifications reduced confidence for motorists in the warning information (Benekohal, 2004; Benekohal and Aycin, 2002, 2004; Benekohal and Rawls, 2004a, 2004b; SRF Consulting Group, 1998). However, unlike the participants in Chugh and Caird's study, the motorists in these field evaluations did not use the displays once they perceived the warning information to be unreliable. Trust in the system, once lost, was not regained.

Other research has investigated the effect of false alarms on operator performance using generic computer-based tracing tasks. Getty, et al. (1995) examined the effect of reliability on the latency of participants' response to warnings by varying the positive predictive value (PPV) of the warning information, that is the probability that a positive indication of the warning truly indicates a dangerous condition. All warning systems have an inherent sensitivity towards detecting an error condition, but whether a warning is issued is based on the amount of evidence a system requires. To understand how PPV can be calculated, it is worthwhile to first consider the predictive value of the warning information with respect to the state of the world. This is shown in Table 2.

Table 2. Predictive Value of Warning Information.

		Truth: Error is Present	Truth: No Error	
Warning Response	**Positive (Warning)**	True Positive	False Positive	*True Positive + False Positive*
	Negative (No warning)	True Negative	False Negative	*True Negative + False Negative*

Table 2 notes two possible "truths" (an error is present or no error is present) and two possible warning responses (a warning is presented or no warning is presented). The PPV rate is defined as the inverse of the true-positive proportion, as described by the formula below:

$$PPV = \frac{numberofTruePositives}{numberofTruePositives + numberofFalsePositives}$$

In Getty, et al., the PPV of the warning was set at one of five levels: 0.25, 0.39, 0.50, 0.61, and 0.75. The task required participants to track a target while simultaneously responding to warnings. Bonuses were provided for good performance on both tasks (i.e., accurate tracking and quick response to true alarms) and penalties were deducted for poor performance.

The results demonstrated that despite the prevalence of high false alarm rates, operators matched their behavior to that of warning reliability. Participants responded slowly to warnings when the PPV was at its lowest (0.25) but responded quickly when the PPV was 0.50 and greater. Feedback regarding the truth of the warning allowed participants to optimize their performance by adopting different strategies in response to system behavior. No one PPV value was determined to lead to "optimal" performance; instead, Getty, et al. noted that the "optimal" value for a given system would depend on costs and benefits specific to that system, and that the costs and benefits would vary as the operating conditions changed. Because the task used by Getty, et al. was abstract, it remains to be investigated whether their findings generalize to predict motorist behavior when approaching an actively protected rail-highway grade crossing.

Although participants in the study conducted by Getty, et al. responded slower to less credible warnings, Bliss, Gilson, and Deaton (1995) found that warning unreliability reduced the likelihood of any response at all. In their study, participants performed a complex cognitive task while simultaneously responding to alarms. Participants were told to expect a 25 percent, 50 percent, or 75 percent probability that a single warning was reliable without any means for verifying that information. Upon presentation of an alarm, participants had 15 seconds to respond. The results showed that the frequency with which participants responded to alarms increased as the reliability of the alarm increased. That is, participants matched their response frequencies to the expected probability of the warning. In fact, the data shows that participants tended to overmatch, responding more frequently than the alarm reliability warranted. Some participants were characterized as "extreme responders." To maximize their success rate, these participants responded 100 percent of the time if the reliability was high (75 percent) and never responded when the reliability was low (25 percent). In general, participants' responses were made in direct proportion to the frequency of the reinforcement, thereby disregarding optimal warning responses behavior from the standpoint of success (e.g., safety).

The results by Getty, et al. (1995) and Bliss, et al. (1995) highlight the importance of reliability and its influence on trust in the warning system, such that if operators do not trust the warning, their responses will be slower and less frequent. Both studies demonstrate consistent shifts in human responses associated with imperfectly reliable warnings that parallel pattern matching and probability learning. In all the studies discussed in this section, the reliability of the warning information was manipulated through the inclusion of false alarms—the presentation of an alarm when no warning state existed. A second failure type is a warning system miss, which is when the warning system does not present an alarm when a warning state exists. In the grade crossing situation, an example of a warning system miss is when the flashing lights and automatic gates fail to activate (and the gate arms remain upright) when a train is approaching a crossing. False alarms and misses exert different effects on behavior, as discussed in the next section.

1.3.2 Warning System Misses

Warning system misses tend to increase operators' vigilance towards potential dangers. For example, in the study by Chugh and Caird (1999), discussed above, participants responded faster to alerts from the in-vehicle warning system following a miss, whereas they responded slower to alerts following a false alarm. Regardless of the failure type, however, participants' trust in the system was consequently reduced .

Meyer (2001) and his colleagues (Cotté, Meyer, and Coughlin, 2001; Maltz and Meyer, 2001) addressed the different effects on false alarms and misses on operators' response to warnings by examining their impact on *compliance* and *reliance*. *Compliance* is observed when the operator responds to warnings that indicate a malfunction (i.e., the operator believes that there is a problem), whereas *reliance* is observed when the operator assumes that no problem exists when no alert is given. The results of a study by Cotté, et al. (2001) indicated that compliance was moderated by the rate of false alarms, whereas reliance was moderated by the rate of misses. In the study, participants performed a collision-avoidance driving task with the assistance of an in-vehicle warning system. The reliability of the warning system was varied at one of two levels: one with a 41 percent rate of false alarms and a 10 percent chance of a miss, and the other with a 10 percent chance of false alarms but a 41 percent chance of a miss. Participants complied with the system when it had few false alarms but relied on the system when the miss rate was low (even though the false alarm rate was high). In two subsequent studies, Meyer (2001) and Maltz and Meyer (2001) found further evidence that reliance was reduced by warning system misses and formed through experience and training.

Dixon, Wickens and Chang (2004) calibrated the presentation of false alarms and misses to examine the attention consequences of the two on-dual task performance. Pilots flew an unmanned aerial vehicle on a military reconnaissance mission in four conditions: no alerting, 100% reliable alerting a 67 percent reliable system with automation false alarms and a 67 percent reliable system with automation misses. The results showed benefits to the presentation of perfectly reliable alerts, but partially reliable alerts reduced performance relative to the baseline condition or worse. Automation false alarms and misses harmed performance in qualitatively different ways, with false alarm prone automation having more detrimental effects on performance than miss prone automation. Misses by the alerting system increased response time to detecting system failures during high workload conditions. On the other hand, the presentation of false alarms decreased the overall detection of system failures. Dixon and Wickens hypothesized that when pilots were interacting with a miss-prone warning system, pilots would occasionally check for targets in the absence of an alarm, but their ability to monitor the situation decreased as workload in the flying task increased. However, when pilots were interacting with a warning system prone to false alarms, they assumed that the presentation of an alarm was in error and did not check the underlying information, thus missing the system failure.

Maltz and Shinar (2004) found similar effects of false alarms and misses on motorist performance when they varied the reliability of warnings presented by an in-vehicle collision avoidance warning system. The warning system presented an alert when the headway between the participants' vehicle and a lead vehicle was too short. The system reliability was manipulated so that in one condition, the reliability was determined by the false alarm rate, and in a second condition, it was determined by the miss rate. In general, the presentation of alerts improved driving performance; motorists who received alerts spent less time in the "danger zone" (i.e., with a headway less than 2 seconds behind the lead vehicle). Interestingly, decreasing the reliability of the warning system did not affect motorists' likelihood of responding to the alert. However, participants made more errors when false alarms were presented by slowing down unnecessarily. Large numbers of missed alerts did not have significant impact on motorists.

Thus, the research shows how warning reliability is reduced by the presence of false alarms and missed signals. The purpose of the present research is to investigate the impact of reduced warning reliability on motorist compliance behavior with active warning devices. We wanted to determine whether reduced warning reliability affects compliance behavior in a predictable way, and if so, if there is a point where warning reliability significantly affects compliance.

1.4 Summary

Noncompliance at grade crossings is a significant safety issue. At active grade crossings, flashing lights or flashing lights and gates warn the motorist of an approaching train. In some cases, noncompliance results from error; for example, the motorist does not detect the traffic control device or does not know what action is required. In other instances, if the motorist is approaching the crossing at the onset of the warning, he/she may determine that it is not safe to stop. Of concern is willful noncompliance. That is, when a motorist consciously and deliberately ignores activated warning signals (flashing lights) and drives around the lowered gates.

The results of observations of motorists at grade crossings suggest that improving the credibility of the warning signal may encourage compliance (Abraham, et al., 1998; Wilde, et al., 1987). Motorists' perception of the credibility of the warning signal may be enhanced by reducing the waiting time at the crossing (e.g., by implementing devices) or by improving the warning reliability. Because warnings are not 100 percent accurate, reduced reliability could increase noncompliance to active warning devices at grade crossings.

The results of the warning literature show that reduced warning reliability has a systematic and predictable effect on operator behavior. Operators use their knowledge of system reliability and prior experience to calibrate their responses to warning systems and respond slower and less frequently to warnings presented by systems that they perceive to be unreliable (e.g., those having a low PPV rate) (Bliss, et al., 1995; Getty, et al., 1995). Alarm false alarms and misses impact operator reliance on the warning system in different ways. False alarms tend to reduce compliance with the warning signal whereas misses tend to reduce reliance (Cotté, et al. 2001; Maltz and Meyer, 2001; Meyer, 2001).

Two studies were conducted to examine the effect of reduced warning reliability on motorists' compliance to flashing lights and gates at grade crossings. We examined two ways in which poor warning reliability manifests itself. Experiment 1 addressed the impact of false alarms on responses to warnings and Experiment 2 focused on the impact of false alarms and missed signals. The PPV of the warning was set at various levels, and signal detection theory was used to measure the impact of changes in PPV on compliance.

2. Experiment 1: Static Task

Experiment 1 examined motorist behavior in response to warning false alarms (i.e., the presentation of a warning when no train was approaching). We evaluated participants' response to eight levels of warning reliability. Participants viewed a series of static images of actively protected highway railroad grade crossings and selected one of two possible response options (Stop or Proceed) to indicate what they would do. Performance feedback was provided once a response was made.

2.1 Method

2.1.1 Participants

Ten John A. Volpe National Transportation Systems Center (Volpe Center) employees were recruited via a center-wide e-mail announcement to participate in this study. Data obtained from two participants (one female and one male) were excluded after learning that they did not follow the instructions. The eight remaining participants were between 21 and 56 years of age (M = 30.3, SD = 14.3). Participants received no payment for their participation. Instead, the experiment took place during working hours in place of normal duties.

2.1.2 Experimental Design

We used a within-subjects experimental design with one independent variable—the PPV rate. Eight PPV rates, representing warning reliability, were examined: .23, .30, .40, .60, .70, .77, .87, and .97. These reliability values were selected based on those used in previous research (Bliss et al. 1995; Getty et al. 1995) and earlier pilot testing that measured the impact of different PPV values on participants' sensitivity and response bias. Two dependent variables measured warning performance: the proportion of valid stops and the proportion of false stops.

2.1.3 Apparatus and Stimuli

The experimental platform was a Pentium IV Dell GX260 computer equipped with a 19-inch flat screen, color monitor, and a standard size keyboard. The experimental trials were generated, displayed, and scored using the software tool, E-prime™ (Version 1.0). Stimuli for the trials consisted of one of five different images of active grade crossings. An example is shown in Figure 1. All images showed an activated grade crossing warning device with the gate in the lowered position and the red light on.

Figure 1. Example Grade Crossing Image

2.1.4 Procedure

Participants read the information sheet, study instructions, signed an authorization form, and completed a demographic questionnaire. Participants were informed about the probabilistic nature of the task, and instructed to base their responses on feedback information provided after each trial. The feedback indicated the accuracy of the last response, which could be used to calculate the overall probability of the accuracy or failure of the warning signal on subsequent trials based on previous trial outcomes.

Participants received 20 practice trials prior to the onset of the experimental period to familiarize participants with the task and its pace. Each trial consisted of four static images:

- a fixation point,
- a grade crossing scene (e.g., see Figure 1),
- a response screen, and
- performance feedback.

The trial began with a screen showing a fixation point (X) that was displayed for 800 milliseconds. This screen was followed by an image of an activated warning device displayed for 500 milliseconds. Next, a response screen appeared that showed two response options: *stop* and *proceed*. Participants based their response on their knowledge of the PPV rate and feedback from previous trials, which could be used to determine the likelihood that the warning device was accurate in the current trial. The response screen remained on the display until participants made their response. Participants responded by pressing the left arrow key (←) to indicate a stop response, and the right arrow key (→) to indicate a proceed response. The responses were mapped to match the vehicle controls (e.g., the brake pedal was positioned on the left, while the gas pedal was positioned on the right). Once participants made their response, a feedback message appeared indicating whether the response was correct or incorrect. This message was displayed for 1000 milliseconds.

Participants were presented with each of the eight PPV rate conditions. For each PPV rate, participants completed five blocks of 60 trials (300 trials total). Participants were instructed to take short breaks between blocks to prevent fatigue. The presentation of both the practice and experimental trials were fully randomized. Data was collected in three, 1-hour long sessions, scheduled over the course of 3 days. Participants were debriefed at the conclusion of the last session and were asked not to discuss this research with coworkers to prevent biasing the participant pool.

2.2 Results and Discussion

In the static task, participants responded to warnings of varying reliability. The stimulus event (a visual warning that a train was approaching) was determined probabilistically, and participants' accuracy was defined to be the proportion of correct responses. The existence of a dangerous condition was determined by the reliability of the warning signal, which indicated one of two warning reliability states: *proper activation* and *false activation*. A proper activation was defined to be a trial when the warning system reliably alerted the motorist about an imminent train arrival. A false activation was defined to be a trial when the warning system provided unreliable information to the motorist by signaling an approaching train when none was present. Participants made one of two possible responses: stop or proceed. Table 3 illustrates the two warning reliability states with respect to participants' response options. The state of the warning (true or false) is shown at the top of the table, and participants' response options (stop or proceed) are shown on the left. The cells in the matrix capture all response outcome categories. Operational definitions for the terms in each cell are provided below the figure.

Table 3. Participants' Response Options to Reliable (proper activation) and Unreliable (false activation) Warnings

		Warning Reliability (Truth)	
		Proper Activation `Presence of a Reliable (True) Warning Signal	**False Activation** Presence of an Unreliable (False) Warning Signal
Participants' Response Options	**Stop** (Compliant)	**Valid Stop** (Compliant and Necessary)	**False Stop** (Compliant but Unnecessary)
	Proceed (Noncompliant Violations)	**High-Risk Violation** (Proceed at High Risk)	**No-Risk Violation** (Proceed at No Risk)

- *Valid Stop*. A valid stop response was analogous to the situation at a grade crossing where a motorist complies with an activated warning signal, brings the vehicle to a complete stop, and awaits the train's arrival. Once the train clears the grade crossing, the gate arm is raised, and the motorist then proceeds safely through the grade crossing. In this study, a valid stop response was recorded when a participant indicated a stop response when a reliable warning was presented. The term "valid stop" was adopted from Raslear's (1996) work on motorist behavior at grade crossings.

- *High-Risk Violation*. A high-risk violation response was analogous to a situation at a grade crossing where a motorist fails to comply with an activated warning signal and violates it by driving around lowered gates. In this instance of gate violation behavior, a vehicle-train collision was imminent because a train was about to enter the grade crossing. The possibility of an accident or a near miss is a consequence associated with this type of gate violation. In this study, a high-risk violation response occurred when a participant gave a proceed response when a reliable warning was presented. This violation type was termed high risk because the motorist proceeded around lowered gates at a risk of colliding with a train.

- *False Stop*. A false stop response was analogous to a valid stop response, where a motorist complies with an activated warning signal, brings the vehicle to a complete stop, and awaits train arrival, but unlike a valid stop, the train never arrives because the warning signal was unreliable. Compliant behavior at actively protected grade crossings generally requires the motorist to stop, and remain stopped until the gate has been raised. In practice, when motorists experience unreliable warning signals, the credibility of the warning system is compromised and motorists may fail to comply with required actions in the future. In this study, a false stop response occurred when a participant gave a stop response to an unreliable warning. The term "false stop" was adopted from Raslear's (1996) work on motorist behavior at grade crossings.

- *No-Risk Violation*. A no-risk violation was analogous to the situation at a grade crossing where a motorist fails to comply (stop) with an activated warning signal, but unbeknownst to the motorist, the warning information is unreliable, so train arrival is not imminent. In the current study, a no-risk violation response occurred when a participant gave a proceed response to an unreliable warning. This violation type is termed no-risk because there is no danger of colliding with a train, even though the motorist proceeds around lowered gates. However, this type of behavior could lead to future high-risk violations, if the driver continues to ignore the information provided by the activated warning signal.

We analyzed data obtained from the static task using two different approaches. The first was a signal detection analysis, which allowed us to examine participants' sensitivity and response bias as a function of the PPV rate. The second was a proportional analysis to determine whether PPV rate influenced compliance in a systematic way. Each analysis allowed us to examine data from a different perspective. The next section provides a brief overview of signal detection theory and presents the results of the signal detection analysis. This is followed by the results of the proportional analysis.

2.2.1 Signal Detection Theory and Analysis

2.2.1.1 Background

Signal detection theory (SDT) provides a framework for studying decisions made in ambiguous situations. This approach to decisionmaking involves the use of a discrete choice task to model an operator's capacity to detect a *signal* against a background of *noise* (Egan, 1975; Green and Swets, 1974). The premise of the paradigm is that there are two states of the world (signal and noise) and two possible human responses ("I detect a signal" versus "I do not detect a signal"), as illustrated in Table 4 (Green and Swets, 1966). Events can be categorized in a 2 X 2 stimulus-response matrix with four possible outcomes: a hit, miss, false alarm, or correct rejection. If a signal is present and it is detected by the operator (a "yes" response), then the response is a *hit*. If a signal is present but is not detected by the operator (a "no" response), then the response is categorized as a *miss*. If no signal is present (i.e., noise) but the operator responds "yes" (i.e., "I detect a signal"), then the response is a *false alarm*. A response of "no" is a *correct rejection*.

Table 4. Four Possible Outcomes in a SDT Matrix

		State of the World	
		Signal	**Noise**
Operator Response	**Yes**	Hit	False Alarm
	No	Miss	Correct Rejection

In SDT, the hit rate is defined as the proportion of "yes" responses to a signal with respect to the number of signal trials. The false alarm rate is defined as the proportion of "yes" responses when no signal is present. The sum of the proportions for the cells in each column in Table 4 will total to one. Thus, although there are four outcome categories, one does not need all four to describe the observer's behavior because the miss and correct rejection rates can be calculated based on the hit and false alarm rates. To illustrate, the miss rate = 1 – hit rate; and correct rejection rate = 1 – false alarm rate (Wickens, 2002).

In addition to outcome categories, SDT makes use of two probability distributions, one that represents the *background noise* and the other representing the *signal*. Figure 2 shows the relationships between the observer's response (outcome categories) within these two hypothetical distributions. The distribution on the right represents the probability of a signal, and the distribution on the left represents the probability of noise. The observer's ability to discriminate between signal and noise is reflected by the amount of overlap between the two distributions.

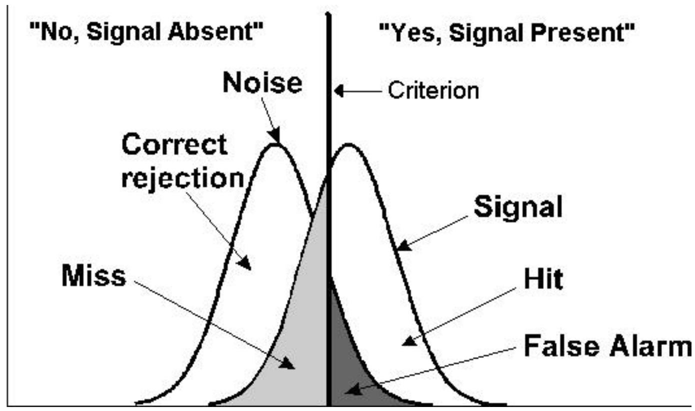


Figure 2. Observer's Responses and Signal and Noise Probability Distributions

SDT isolates the inherent detectability of the signal from attitudinal or motivational variables that influence the observer's criteria for judgment. Signal detection theory provides a measure of two processes: the observer's perceptual *sensitivity*, and his/her *response bias*. Sensitivity (i.e., detectability) measures an operator's ability to discriminate between signal and noise. The symbol *d´* is a widely used sensitivity measure, which expresses the distance between the means of the two distributions in standard deviations. In Figure 2, sensitivity is represented by the distance between the means of the two distributions. Computations of this parametric measure is based on standardized z-scores (*Z*) of hit and false alarm distributions, where,

$$d' = Z\text{ (hit rate)} - Z\text{ (false alarm rate)} = \frac{\overline{x_S} - \overline{x_N}}{SD_N},$$

where $\overline{x_S}$ = the mean of the signal distribution,

$\overline{x_N}$ = the mean of the noise distribution, and

SD_N = the standard deviation of the noise distribution

When d´ = 0, the operator does not distinguish between signal and noise. As d´ increases, it indicates that the operator is better able to distinguish between the two. A sensitivity measure can be summarized by its receiver operating characteristic (ROC) curve that graphically depicts trade-offs between hits and false alarm rates, thereby providing a visual representation of decision behavior. The ROC space is shown in Figure 3.

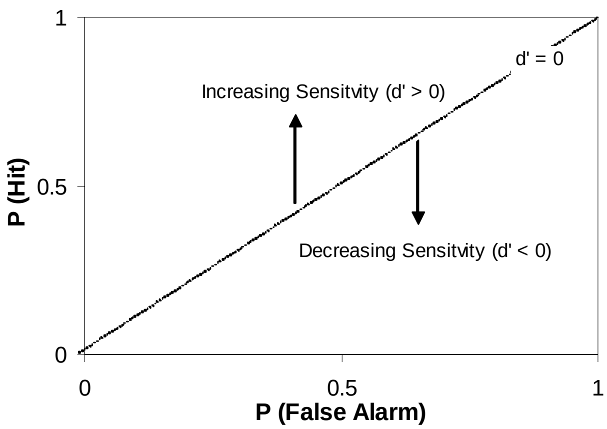


Figure 3. ROC space.

The diagonal line through the center of the ROC space indicates the case where d′ = 0; that is, the operator is not sensitive to the presence/absence of the signal. As d′ increases, sensitivity shifts to the left as the proportion of hits to false alarms increases, indicating better signal detection performance.

Response bias (λ) represents an operator's willingness to indicate "yes, there was a signal" or "no, there was no signal". Response bias represents an observer's decision criterion and indicates how much evidence is needed by the operator before s/he can conclude that a signal was present. According to Wickens (2002), response bias can be estimated by lambda-center (λ_c), which takes into account both hit and false alarm rate and therefore, considers both signal and noise distributions. The value of λ_c can be found from d′ and λ using the following formula:

$$\lambda_c = \lambda - 0.5\, d', \text{ where } \lambda = -Z \text{ (false alarm rate)}$$

Substituting the equation for d′, provided above, the calculation for λ_c is equal to:

$$\lambda_c = -\tfrac{1}{2}\, [Z \text{ (false alarm rate} + Z \text{ (hit rate)}]$$

The response criterion is represented by the solid line in Figure 2. The observer is performing ideally when $\lambda_c = 0$. Changes in the response criterion cause a shift to the left or right. When λ_c is positive (i.e., a shift to the right), the observer is exhibiting conservative behavior. The response criterion shifts closer to the signal distribution, indicating a bias to say "no, there is no signal" and to proceed. This behavior leads to few hits but many correct rejections. On the other hand, when λ_c is negative (i.e., a shift to the left), the observer is exhibiting liberal behavior. The response criterion shifts closer to the noise distribution, indicating a bias to say "yes, there is a signal" and to stop. This shift results in a higher hit rate but lower correct rejection rate (Wickens, 2002). Response bias can also be estimated from the ROC curve, such that data points that fall in the upper, right corner of the ROC curve indicate a bias to say "yes, there is a signal," while points that fall in the lower left corner indicate a bias to say "no, there is no signal" (Raslear, 1996).

SDT is broadly applied in human factors research (e.g., Dow, Thomas, and Johnson, 1999; Maltz and Shinar, 2000; Multer, Conti, and Sheridan, 2000; Nakata and Noel, 2001; Peterson, Uhlarik, Raddatz, and Ward, 1999). Given its strength in describing human detection performance, SDT is increasingly used in research on warnings (e.g., Cotté, Meyer, and Coughlin, 2001; Jurgensohn, et al., 2001; Maltz and Meyer, 2001; Meyer, 2001, 2004; Meyer and Bitan, 2002). In fact, Raslear (1996) applied SDT to study motorist compliance with warning systems at highway-railroad grade crossings. In his paper, Raslear compared the response of a motorist at a grade crossing with an approaching train to that of an operator attempting to detect a signal in background noise. Table 5 depicts the signal detection framework applied to the task of the motorist approaching a highway-railroad grade crossing.

Table 5. Signal-Response Matrix for a Motorist at a Grade Crossing (from Raslear, 1996)

		State of the World	
		Train is close	**Train is not close**
Motorist Response	**Yes (Stop)**	**Valid Stop** (motorist stops at crossing)	**False Stop** (motorist stops unnecessarily)
	No (Proceed)	**Accident** (motorist doesn't stop)	**Correct Crossing** (motorist crosses tracks safely)

We applied the signal-response framework, shown in Table 5, to the data collected in the present study. The rate of valid stops, false stops, accidents, and correct crossings were calculated to measure participants' sensitivity (d´) and response bias (λ_c), using the two formulas above. We then examined the changes in sensitivity and response bias as a function of the PPV rate to determine the effects of signal reliability on participants' decisions to stop or proceed.

2.2.1.2 Sensitivity

Figure 4 illustrates the changes in mean sensitivity (d´; i.e., the ability to differentiate between proper and false warning activations) as a function of the PPV rate (i.e., the reliability of the warning system). The calculation of d´ was based on the probability of hits and false alarms (the top two cells in the SDT matrix in Table 3). A repeated measures Analysis of Variance (ANOVA), conducted with PPV as a within-subjects factor and the sensitivity values as the dependent variable, indicated a significant effect of PPV, $F(7,49) = 10.3$, $p < .001$. This is shown in Figure 4. Polynomial contrasts indicated a significant linear effect with means decreasing across PPV condition, $F(1, 7) = 34.9$, $p < .001$. *T*-tests were performed to evaluate differences between the PPV condition means. Only one difference was significant, that between PPV = .87 ($M = .12$, $SD = .48$) and PPV = .97 ($M = .90$, $SD = .42$), $t(7) = -2.95$, $p < .05$. This finding suggests that as warning system reliability decreased from PPV = .97 to PPV = .87, so did participants' ability to differentiate between proper warning activations and false warning activations.

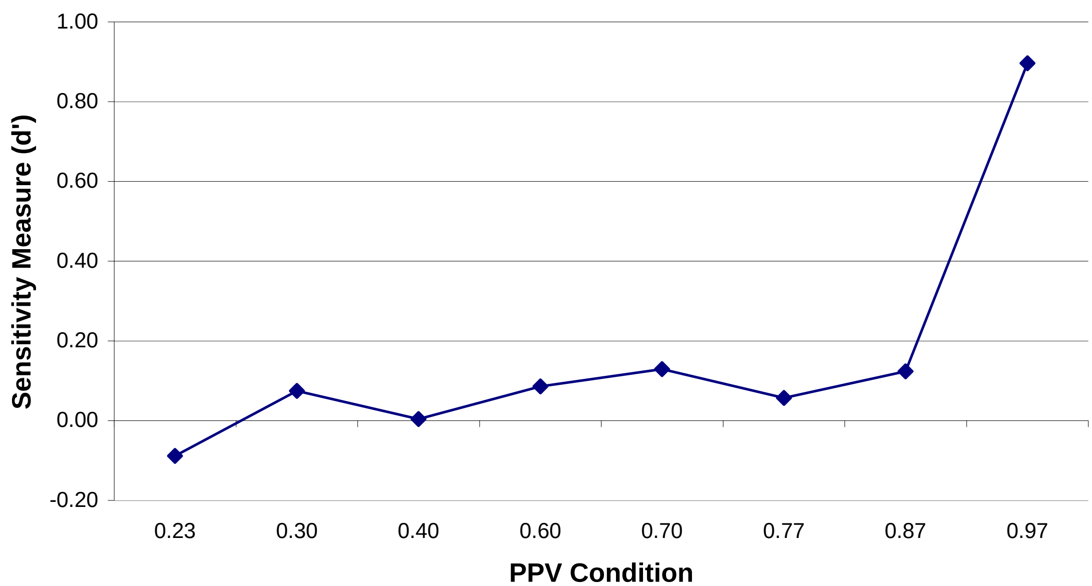


Figure 4. Mean Sensitivity as a Function of Warning System Reliability (PPV)

Another way to illustrate the sensitivity differences across PPV conditions is to graph the probability of hits as a function of the probability of false alarms for each PPV condition as a ROC curve. This is shown in Figure 5, which shows eight ROC points, one per PPV condition. The diagonal line spanning the chart represents chance performance. As sensitivity improves (i.e., as the motorist is better able to detect whether the warning is reliable or unreliable), the points along the curve would move from the center of the chart to the upper left corner. However, as Figure 5 shows, sensitivity to the PPV rate approximated chance level performance for all warning system reliability levels, except for PPV = .97. The performance trend shown by the ROC curve reflects participant's *in*ability to differentiate between reliable and unreliable warnings at PPV rates below 0.97 and is consistent with that shown in Figure 4.

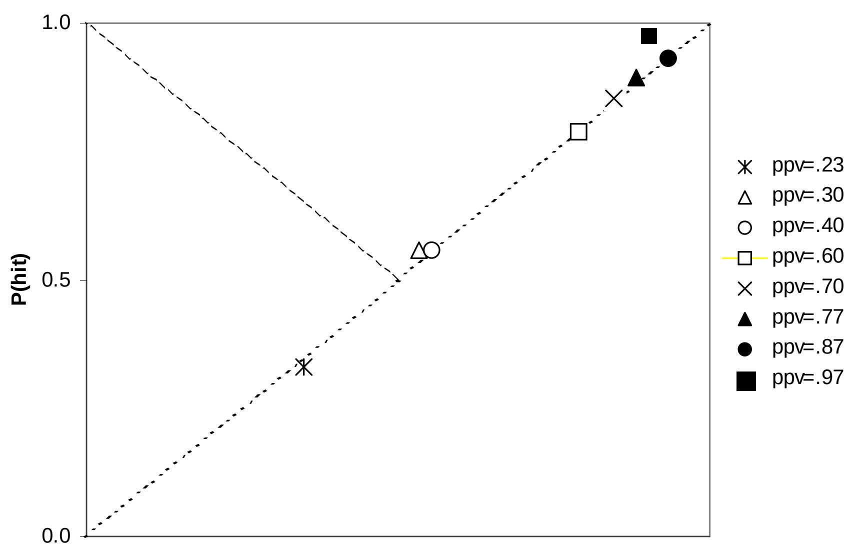


Figure 5. ROC by PPV.

2.2.1.3 Decision Criterion (Response Bias)

Figure 6 illustrates the changes in mean response bias (λ_c; i.e., the willingness to choose whether to stop or proceed) as a function of warning system reliability, the PPV rate. A within-subjects repeated measures ANOVA indicated a significant effect of PPV on response bias, $F(7, 49) = 17.2$, $p < .001$. Polynomial contrasts showed a significant linear effect with means decreasing as the PPV rate increased, $F(1, 7) = 52.9$, $p < .001$. In other words, as warning system reliability decreased, participants' response bias shifted to the right; consequently, participants showed an increased bias towards proceeding. One-paired sample t-tests were performed based on the graphical results to evaluate differences between the PPV conditions means. Only one difference was significant: PPV = .40 (*M* = -.32, *SD* = 1.02) and PPV = .60 (*M* = -1.18, *SD* = 1.03), $t(7) = 3.13$, $p < .05$.

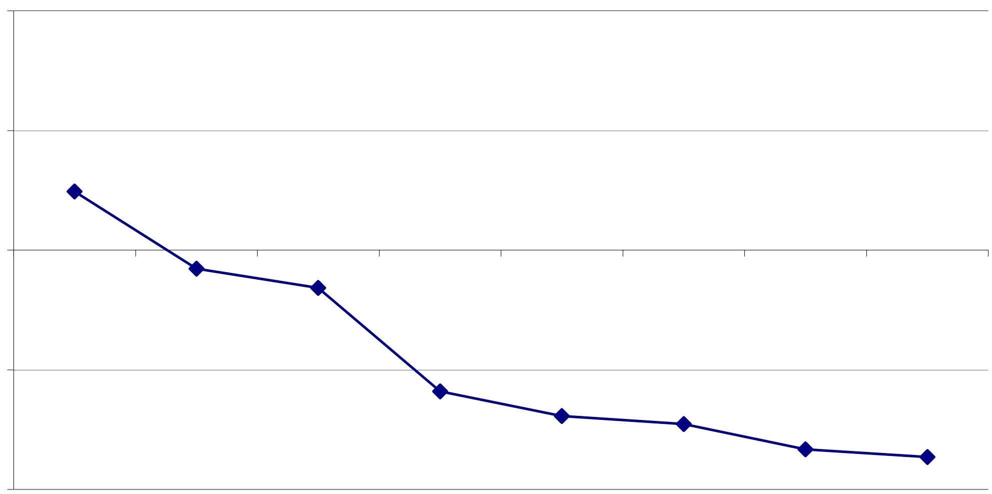


Figure 6. Mean Response Bias as a Function of Warning System Reliability (PPV)

2.2.2 Analysis of Response Proportions

We assessed participants' likelihood of compliance by examining the proportion of valid stops (hits) and false stops (false alarms; see Table 3). The proportion of valid stops was calculated by dividing the total number of valid stop responses by the total number of trials on which the warning system was properly activated. The proportion of false stops was calculated by dividing the total number of false stop responses by the total number of trials on which the warning system was falsely activated.

Whereas we chose to examine motorist behavior in terms of compliance, it can also be considered by observing rates of noncompliance. Measures of noncompliance are the proportion of high-risk violations, calculated by subtracting the valid stop proportion from 1, or no-risk violations, calculated by subtracting the proportion of false stops from 1. Since we chose to focus on measures of compliance, no analysis was conducted on the data for high-risk and no-risk violations. However, these data and a discussion of these measures are provided in Appendix A.

2.2.2.1 Compliance with Properly Activated Warning Systems

Compliance was defined as the proportion of valid stops—participants' stopping when the warning system was properly activated. We hypothesized that compliance with properly activated signals, as measured by the proportion of valid stop responses, would increase as the PPV rate increased, and in fact, the data confirms this hypothesis. A within-subjects repeated measures ANOVA assessing valid stop response proportion as a function of the PPV rate showed a significant effect, $F(7,49) = 20.1$, $p < .001$. Polynomial contrasts indicated a linear relationship, PPV, $F(1, 7) = 39.7$, $p < .001$, as shown in Figure 7. The finding corroborates existing warnings research that found performance variability to be a function of warning reliability (e.g., Bliss et al., 1995; Getty et al. 1995; Maltz and Meyer, 2001)

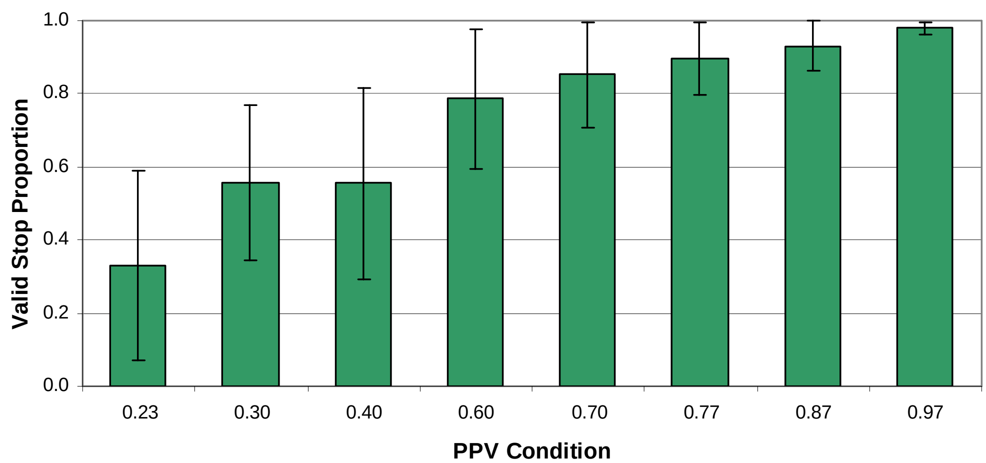


Figure 7. Means and Standard Deviations for Valid Stop Proportion as a Function of Warning System Reliability

Visual inspection of Figure 7 suggests that valid stop response proportions approximated warning system reliability levels across the PPV spectrum. To further examine this relationship, both the valid stop response proportions and PPV reliability levels were graphed simultaneously. This is shown in Figure 8, which provides an illustration of how closely participants 'matched' their responses to warning reliability (PPV). In the figure, the line for "Expected Response Matching" reflects all eight PPV conditions while the line for "Observed Valid Stop Response" reflects participants' compliance with a reliable, or properly activated warning system. The pattern of the two lines resembles probability matching behavior (Bliss, et al., 1995).

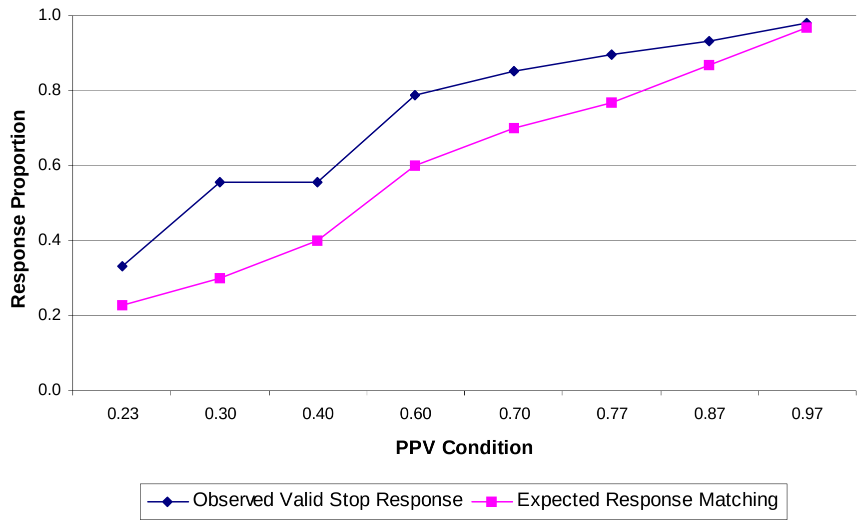


Figure 8. Ideal Pattern Matching and Observed Valid Stop Proportions

As Figure 8 shows, the proportion of valid stop responses was higher than the expected responses based on the PPV rate for each PPV condition. This finding is consistent with the results of Bliss et al. (1995) and demonstrates that participants not only acquired information about the probabilities of the outcomes, but that they also were less risky when making their responses. Despite the evident overmatching, however, the degree of association between valid stop proportions and expected response matching (PPV) was very high. A Pearson Product-Moment correlation coefficient computed among the expected and obtained scores indicated a statistically significant relationship between valid stop proportions and expected response matching, $r(7) = .97$, $p < .01$.

2.2.2.2 Compliance with Falsely Activated Warning Systems

We expected that compliance with falsely activated warnings (i.e., an alarm false alarm), as measured by the proportion of false stop responses, would decrease as the PPV rate increased (i.e., as warning system reliability increased). Figure 9 demonstrates the means and standard deviations for false stop proportions as a function of the PPV rate, and shows that contrary to our hypothesis, the proportion of false stops *increased* as the warning reliability increased. A within-subjects repeated measures ANOVA examining the proportion of false stops as a function of the PPV rate showed a significant effect, F (7, 49) = 20.1, $p < .001$. Polynomial contrasts also showed a significant linear effect, $F(1,7) = 39.2$, $p < .001$, indicating a significant increase in false stop proportion as the PPV rate increased, a finding similar to that reported by Maltz and Shinar (2004). Thus, participants regarded performance feedback differently than we expected and overestimated the overall warning reliability. Consequently, they provided an overabundance of stop responses, which in turn increased the valid and false stop rates.

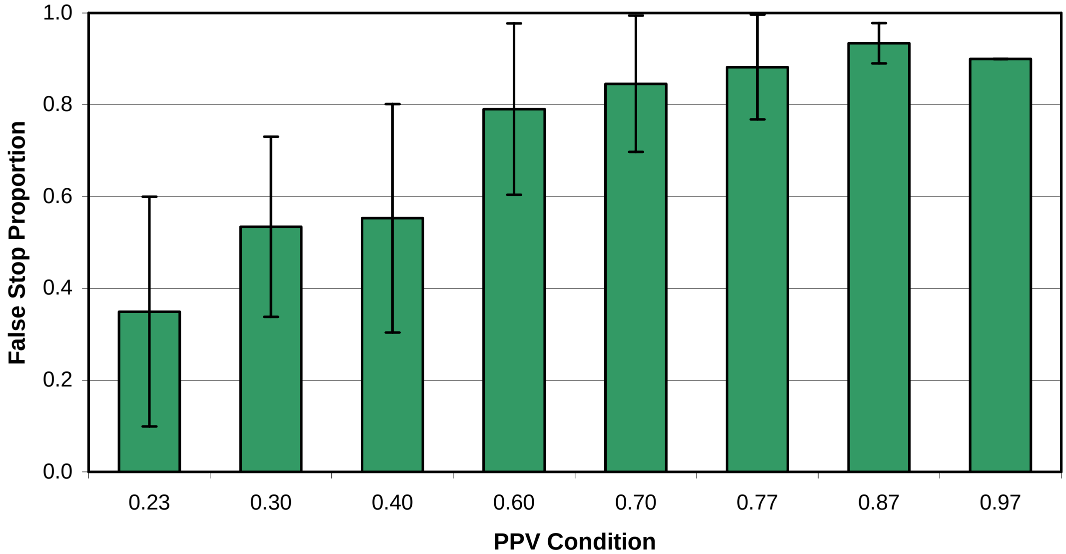


Figure 9. Mean and Standard Deviations for False Stop Proportions as a Function of Warning Reliability (PPV)

2.2.2.3 Overall Compliance

A measure of overall compliance, based on the proportion of valid stop and false stop responses, was obtained by computing a mean of those response proportions. Overall compliance is a measure of participants' total stopping behavior. We expected that overall compliance would increase as the PPV rate increased, and that data confirms this hypothesis, as shown in Figure 10. A within-subjects repeated measures ANOVA examining the effect of PPV on overall compliance revealed a significant effect of PPV rate, $F(7, 49) = 20.49$, $p < .001$. Polynomial contrasts indicated a significant linear effect with means increasing with PPV condition, $F(1, 7) = 39.63$, $p < .001$. This result is consistent with previous research on warnings suggesting that compliance is a function of warning system reliability (e.g., Bliss, et al., 1995; Maltz and Meyer, 2001; Meyer, 2001).

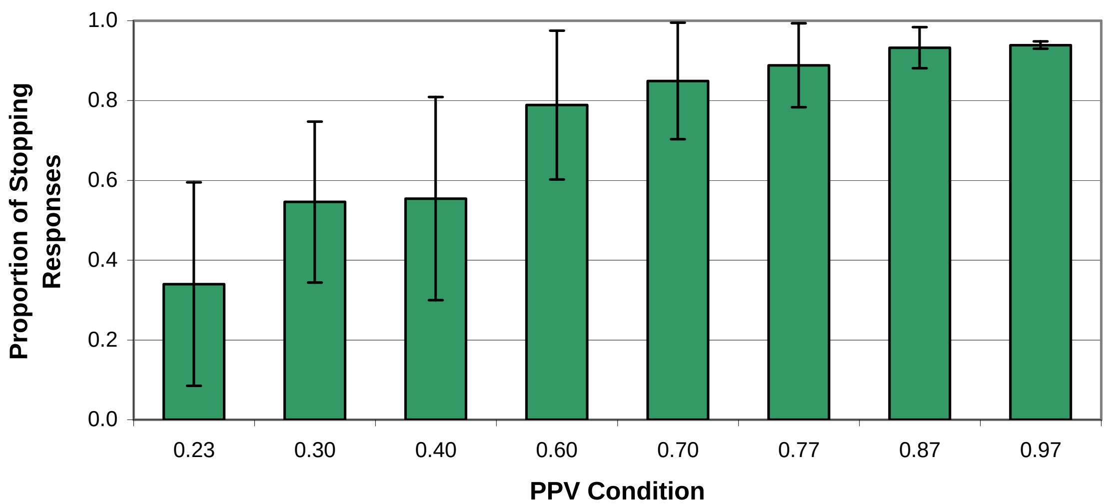


Figure 10. Means and Standard Deviations for Overall Compliance Response Proportions.

2.2.3 Summary of Findings

The results of the signal detection analysis and the examination of response proportions both showed that decreasing the PPV rate had the predictable effect of reducing compliance. However, participants were not sensitive to changes in the PPV rate, unless the PPV was high. Analysis of sensitivity (d′) revealed that participants were able to differentiate between proper warning activations and false warning activations only when the PPV rate changed from .87 to .97. That is, a high degree of system reliability (at least .87) was needed for participants to differentiate correctly between warning signals that were properly activated from those that were falsely activated. This finding has critical implications. If motorists perceive or decide that a warning system is unreliable due to a high number of previous false alarms, they may engage in violating the lowered gates because they do not believe train arrival is imminent. This behavior can lead to an accident or to a near miss when the warning signal is not in error.

Similarly, the analysis of response bias (λ_c) revealed that participants were more likely to proceed than stop when faced with an ambiguous grade crossing situation with a low reliability signal (i.e., a low PPV rate). A significant increase in the response bias (i.e., a shift to the right) was found when the PPV rate decreased from .60 to .40, suggesting that as the warning system became less reliable, participants exhibited riskier behavior. In other words, when participants perceived warnings to be unreliable, they were more likely to proceed through the grade crossing and violate lowered gates. On the other hand, when the participants perceived the system to be reliable (at a PPV rate equal to or greater than 0.60), they exhibited a bias towards stopping and complying with the signal.

Results of the proportional analysis examining compliance with a properly activated warning system revealed that the PPV rate affected valid stop and false stop proportions systematically. In first considering the results for valid stops, the data showed that the proportion of valid stop responses increased as the warning system reliability increased. This finding is consistent with the hypothesis set forth in Lerner et al. (1990), suggesting that warning reliability impacts compliance behavior at grade crossings. In our study, participants tended to match their responses to the predictive ability of the warning system as defined by the PPV rate. This finding is consistent with that reported by Bliss et al. (1995), and Getty et al. (1995). In fact, participants adjusted their responses so that the degree of association between valid stop proportion and predictive ability, or expected matching (PPV), was very high (see Figure 8).

Because participants exhibited compliant behavior with reliable warning systems, as warning system reliability increased, so did participants' false stopping responses. From a traffic law standpoint, this behavior seems desirable. However, from a human factors perspective, these false stops contribute to motorist frustration at grade crossings and facilitate mistrust in the warning system. Consequently, in the future, motorists may distrust an accurate active warning signal and engage in gate violation behavior, leading to an accident or near miss.

Collectively, the proportion of valid stops and false stops, used as a measure of overall compliance, revealed that overall stopping behavior increased as a function of warning system reliability. The findings are consistent with the results of the literature examining operator responses to warnings that find consistent shifts in behavior because of system unreliability, manipulated here by the PPV rate (e.g., Bliss, et al., 1995; Getty, et al., 1995). The results support the hypothesis set forth by Lerner et al. (1990) that motorists' perception of warning system reliability could affect compliance with active warning devices at grade crossings, and that motorists would be less likely to comply with what they perceived as an unreliable warning.

One limitation of Experiment 1 is that it examined motorist behavior using a static environment. Participants were shown images of active grade crossings and asked to indicate their response (stop or proceed), and there was no consideration of other motivations (e.g., time pressure) that might factor into a motorist's decision to comply at a grade crossing that would be typical of a realistic driving environment. Additionally, the current study focused only on the impact of false alarms on motorist behavior and did not examine the effect of missed events. As Meyer and his colleagues have reported, false alarms and misses have different influences on operator behavior (Cotté, Meyer, and Coughlin, 2001; Maltz and Meyer, 2001; Meyer, 2001). Thus, Experiment 2 sought to examine the effect of system unreliability in a more realistic dynamic driving environment and to evaluate the impact of false alarms and missed signals on motorists' decisions at grade crossings.

3. Experiment 2: Driving Task

Experiment 2 examined how motorist responses to grade crossing warning signals were influenced by false alarms (i.e., a warning when no train was approaching) and missed signals (i.e., the failure of the warning system to signal an approaching train). Participants performed two tasks: a *priming task*, similar to the task used in Experiment 1, and a *driving task* in a simulator. The goal of the priming task was to set motorist's expectations about warning system reliability in the driving task. Participants viewed a series of static images of gated highway-railroad grade crossings that showed an active or inactive warning signal and selected one of two possible response options (stop or proceed). The reliability of the warning signal (i.e., the PPV rate) was manipulated by varying the rate of false alarms and misses. In the driving task, participants drove a simulated vehicle through a course with 24 active grade crossings with partially reliable warning systems.

3.1 Method

3.1.1 Participants

Twenty-seven Volpe Center employees were recruited via an email announcement to participate in the study. Data obtained from two of the participants were discarded because they were unable to complete all three experimental sessions. Of the 25 participants who completed all the sessions, 14 (56 percent) were female and 11 (44 percent) were male. Participants were not paid for their time since the experimental periods took place during working hours in place of normal duties. Additionally, participants were presented with an opportunity to obtain three gift certificates (one for completion of each experimental session). The mean age of the participants was 35.1, with a standard deviation of 13.1, and a range from 20 to 61 years. On average, the participants had 12 years of driving experience.

3.1.2 Experimental Design

The experiment used a within-subjects design. Three PPV rates were examined: .40, .60, and .83. The sequence in which participants received the PPV conditions was randomized and counterbalanced across participants. For each PPV rate, participants completed a priming task and a driving task (both tasks are described in more detail in the next section). Participants completed the priming task before the driving task, and they completed both tasks for one PPV rate before moving on to the next PPV rate.

3.1.3 Tasks

Priming Task. The priming task employed the same paradigm as that utilized in Experiment 1 with three modifications. First, images of *in*activated warning signal devices were included in the set of warning images. As a result, participants saw scenes of a grade crossing with gates in the raised position (inactive) and scenes of a grade crossing with gates in the lowered position (active). Second, a train horn sound was incorporated to indicate train arrival. The sounding of the train horn was always reliable (i.e., it always indicated an approaching train). Third, the feedback screen indicating whether participants' performance was correct or incorrect for the trial was eliminated.

Thus, these modifications yielded four different situations:

1. an image of an activated warning signal and the train horn sound,
2. an image of an activated warning signal and no train horn sound,
3. an image of an inactivated warning signal and train horn sound, and
4. an image of an inactivated warning signal and no train horn sound.

Driving Task. The driving task was conducted using a low-fidelity fixed-based driving simulator created with a Direct-X Microsoft platform workstation. The visual image was displayed using a Barco™ projector on an 8 ft by 10 ft wall-mounted screen. Participants were seated 15 ft from the projection screen. Although the driving simulator was capable of simulating motion, no motion-base was available, so the speed and braking profile used by the vehicle, which was modeled after a medium size sedan, may have been somewhat inconsistent with visual cues. The vehicle was controlled with a Logitech Wingman™ force feedback steering wheel and an accelerator and brake pedal that were positioned on the floor. The distance between the vehicle's controls and the seat were adjusted depending on participants' anthropometrics. A wooden box simulated the vehicle's enclosure.

The simulated driving course was 12-miles long with 24 grade crossings. Figure 11 shows the geographical map of the driving course and grade crossing locations. The simulated road environment consisted of a two-lane highway with a 45 miles per hour (mph) posted speed limit (note that the maximum speed at which the vehicle could travel was set at 55 mph).

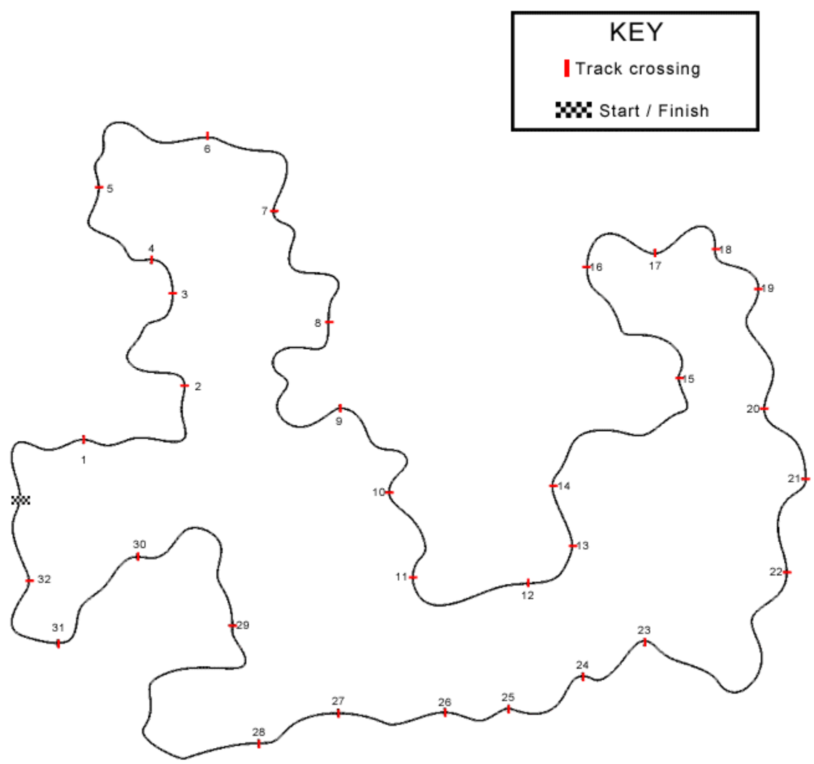


Figure 11. Driving Course

The approach to every highway-rail grade crossing consisted of an advance warning sign, crossbuck, RXR pavement markings, stop line, and automatic gate with flashing lights. Figure 12 shows a scene from a motorist approaching a grade crossing in the driving simulator.

Figure 12. Typical Grade Crossing

The vehicle speedometer was located in the lower left-hand corner of the screen. The road markings, signs, and railroad warning devices presented in the driving scene were designed according to the regulations set forth in the MUTCD (FHWA, 2000). Road markings were provided only in the lane of travel. As illustrated in Figure 12, the gate is lowered and the flashing light is activated. In the driving simulator, the gate took 5 seconds to descend (and ascend), and the red lights alternated when flashing. The design of the simulated grade crossing environment did differ from an actual grade crossing in one important way—the second gate arm was omitted from the design. This modification was made to simplify maneuvering around lowered gate arms in the simulator when participants chose to do so. Additional information about the grade crossing design is listed in Appendix B. Details regarding the reliability of the warning system at each crossing for each PPV rate are provided in Appendix C.

Dependent Measures. Three measures were used to examine motorist behavior in the priming task: compliance, sensitivity, and response bias. Compliance with the warning devices was assessed via the proportion of correct responses (PCR), as calculated from the following formula:

$$PCR = \frac{[P(hit) + P(cr)]}{[P(hit) + P(miss) + P(fa) + P(cr)]}$$

Similar to that used in Experiment 1, a signal detection analysis was used to evaluate participants' sensitivity (d′) to differentiate between reliable and unreliable warnings and their bias to stop or proceed (estimated using lambda center, λ_c).

In the driving task, the compliance, sensitivity, and response bias measures were also used to assess behavior. Additionally, three other descriptive measures were collected to evaluate driving performance; collision frequency, task completion time, and train time to crossing.

Collision frequency was measured by counting the number of vehicle collisions with trains, lowered gates, and roadside objects. Driving task completion time was measured from the time a participant started driving until he/she cleared the last grade crossing. Finally, train time to crossing was measured by the time, in seconds, it took a train to reach the grade crossing after the vehicle crossed the track.

3.1.4 Procedures

Each participant was presented with the experiment instructions and consent form. The documents described the basic experimental protocol and the risks and benefits associated with participation and provided assurance of information confidentiality. Each participant also completed a demographic questionnaire.

Participants completed the priming task first. Each session began with a 20-trial practice session to familiarize participants with the task and its pace. Participants then completed 25 experimental trials. The priming task took about 20 minutes.

Participants then received a two-minute simulator training to familiarize themselves with the vehicle controls and the driving environment for the driving task. During the training period, participants drove for approximately one mile through the driving course and approached the first two grade crossings. This training session was provided prior to each experimental driving session. During the experimental session, participants drove through the test course (shown in Figure 11) and experienced different warning system events at grade crossings depending on the experimental condition.

Participants were presented with the opportunity to earn incentives based on their performance in the driving task. Participants were encouraged to complete the driving task as quickly as possible to earn a gift certificate of one of three different values ($25, $50, and $75). This payoff structure was adopted to simulate more realistic driving conditions in the task (e.g., time pressure and competing motivations).

The data was collected in three separate sessions lasting 3 hours each over a 3-day period. During each session, the participant was exposed to a different PPV rate. Participants were debriefed at the conclusion of the experiment and asked not to discuss this study with fellow co-workers.

3.2 Results and Discussion

3.2.1 Signal Detection Approach

One of the challenging tasks in conducting this research was conceptualizing what the 'signal' is in the grade crossing situation. Is the signal defined by the presence of a train or by the lowered gate? To help answer this question, we looked to traffic law for guidance. The law requires motorists to stop when a gate is in the lowered position and wait until it is raised, regardless of whether a train arrives or not. Because the definition of compliance by law is dependent on the state of the warning device, and not the presence of a train, we defined our signal to be a lowered gate.

An active warning system at a grade crossing is a dynamic sensor-based signaling system. Two possible states of the warning system exist—it is functioning properly, or it is not—and two possible outputs—a warning is activated, or it is not. The two states of the warning system and two states of the warning output create four combinations, which can be defined in terms of what the motorist encounters at a grade crossing:

1) *Warning system proper activation*. The warning system provides reliable information; the warning signal is activated and a train arrives at a crossing;

2) *Warning system proper* in*activation*. The warning system provides reliable information; the warning signal is *not* activated and a train does not arrive at a crossing;

3) *Warning system false activation*. The warning system provides unreliable information; the warning signal is activated but a train does not arrive at a crossing;

4) *Warning system fails to activate*. The warning system provides unreliable information; the warning is *not* activated, but a train arrives at a crossing.

Meyer (2004) showed that a 2^3 (2 x 2 x 2) matrix could be used to describe these four combinations of warning system outputs for each of two possible operator responses (stop or proceed). This is illustrated in Table 6.

Table 6. Extended SDT Matrix.

	WARNING SYSTEM RELIABILITY			
	Warning System Functions Properly		**Warning System Fails**	
	Proper Activation Activated Signal and Train Approaching	**Proper Inactivation** Inactivated Signal and Train Not Approaching	**False Activation** Activated Signal and Train Not Approaching	**Failure to Activate** Inactivated Signal and Train Approaching
Stop	**Valid Stop** (Compliant)	**Inappropriate Stop** (Noncompliant)	**False Stop** (Compliant)	**Fortunate Stop** (Noncompliant)
Proceed	**High Risk Violation** (Noncompliant)	**Proceed Safely** (Compliant)	**No-Risk Violation** (Noncompliant)	**Accident Risk or Proceed at High Risk of an Accident** (Compliant)

Note: Compliance (in parentheses) refers to traffic law.

The responses shown in each cell are characterized below with respect to traffic law definitions of compliant or noncompliant behavior at grade crossings as mediated by the reliability of the warning system. We first consider the four cells in the left side of the matrix in Table 6 that characterize motorist responses when the warning system is functioning properly and then discuss the four cells in the right side of the matrix that describe motorist responses when the warning system fails.

Warning System Functions Properly. The matrix in Table 6 shows two possible warning events at a grade crossing: the warning system presents a signal that correctly indicates the train's arrival (a warning system hit), or the warning system presents no signal and correctly indicates that no train is arriving (a warning system correct rejection). The four cells specify four types of motorist responses to these two warning states: a valid stop, a high-risk violation, an inappropriate stop, and proceed safely. These responses differ in terms of the consequences for the motorist as described below.

- *Valid Stop.* This response represents a situation at an active grade crossing where a warning signal is activated, and the motorist complies with the warning system signal by bringing the vehicle to a complete stop and awaiting the train's arrival (FHWA, 2003). This response type is compliant with traffic law.

- *High-Risk Violation*. This response represents a situation at an active grade crossing where a motorist fails to comply with an activated warning signal and violates it by driving around the lowered gate arms. Because the warning system functions properly, it provides reliable information about an imminent train arrival and a vehicle-train collision may result. This response type is *not* compliant with traffic law.

- *Inappropriate Stop.* This response represents a situation at an active grade crossing where the warning signal is not activated, but the motorist brings the vehicle to a complete stop and looks for an approaching train. This response type is not required and therefore considered *not* compliant with traffic law.

- *Proceed Safely.* This response represents a situation at an active grade crossing where the warning signal in not activated, and the motorist proceeds safely though the grade crossing. This response type is compliant with traffic law.

Warning System Fails. The right side of the matrix in Table 6 shows two possible types of warning malfunctions at a grade crossing: a warning system that incorrectly indicates a train's arrival (a warning system false alarm), and a warning system that fails to indicate that a train's arrival is imminent (a warning system miss). The four cells specify four types of motorist responses to these two failure events: a false stop, a no-risk violation, a fortunate stop, and proceed at a high risk of an accident. These responses differ in terms of consequences for the motorist as described below.

- *False Stop.* This response represents a situation at an active grade crossing where the motorist complies with an activated warning signal by bringing the vehicle to a complete stop and awaiting the train's arrival. However, since the warning signal has malfunctioned, the motorist waits for a train that never arrives. Consequently, the motorist may be frustrated by the wait time, confused by the information received from the warning system, and experience pressure from other motorists to violate the gate. Nevertheless, stopping at the crossing is compliant with traffic law.

- *No-Risk Violation.* This response represents a situation at an active grade crossing where the motorist fails to comply with an activated warning signal and violates it by driving around the lowered gate arms. However, since the warning was falsely activated, no train is approaching. Since a collision with a train will not occur, the risk associated with this violation is low. Nevertheless, this response type is *not* compliant with traffic law.

- *Fortunate Stop.* This response represents a situation at an active grade crossing where a motorist stops when the warning device is not activated (e.g., to look for an approaching train). In this situation, the warning signal has failed to indicate an imminent train arrival (a warning system miss), so by stopping despite the lack of a warning signal, the motorist escapes a potential collision with a train. Although the motorist's cautious response in this case is beneficial, it is *not* compliant with traffic law.

- *Proceed at High Risk of an Accident.* This response represents a situation at an active grade crossing where the warning devices are not activated, and the motorist proceeds through the grade crossing. However, since the warning device fails to indicate an imminent train arrival (a warning system miss), the motorist unknowingly proceeds through the crossing at the great risk of colliding with a train. This response type is compliant with traffic law.

The framework described by the matrix in Table 6 was used to analyze the data obtained in the priming task and SDT measures of the driving task.

3.2.2 Priming Task

Priming task performance was assessed via three dependent variables: compliance (calculated as the proportion of correct responses, PCR), sensitivity (d′) and response bias (λ_c). Details for each participants' performance (i.e., their rate of hits, misses, false alarms, and correct rejections) is provided in Appendix D. Within-subjects repeated measures ANOVA were used to assess changes in all three variables as a function of the PPV rate. The discussion of results begins with an examination of compliance (i.e., PCR). This is followed by the presentation of results regarding participants' sensitivity and response bias.

3.2.2.1 Compliance (PCR) in the Priming Task

On the basis of the results in the first study, we expected that compliance in the priming task would increase as PPV increased. This expectation was confirmed, $F\ (2, 48) = 48.1$, $p < .001$[1]. Figure 13 illustrates changes in PCR as a function of PPV. Paired sample *t*-tests conducted to evaluate differences between the three PPV conditions found significant differences between all three PPV rates at the $p < 0.05$ level (PPV = .40: *M* = .71, *SD* = .14; PPV = .60: *M* = .83, *SD* = .09; PPV = .83, *M* = .92, *SD* = .04).

[1] A Bonferroni correction was used to control for Type I error. The simplest form of the Bonferroni alpha level adjustment starts with the desired family-wise error ($\alpha = .05$) and divides that probability equally among all of the comparisons (Keppel, 1991). The new pair comparison significance level was obtained by dividing the family-wise α level of .05 by the number of comparisons (2), and resulted in a *p* value of .025.

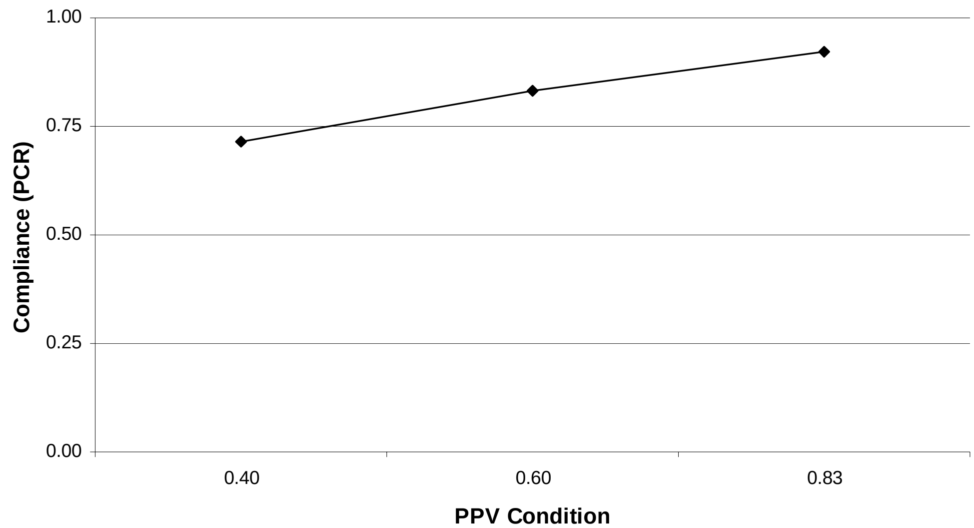


Figure 13. Mean Warning Compliance (PCR) as a Function of Warning System Reliability (PPV) in Priming Task.

3.2.2.2 Sensitivity (d´) in the Priming Task

A within-subjects repeated measures ANOVA revealed an overall significance of PPV rate on sensitivity, F (2, 48) = 24.9, $p < .001$, as illustrated in Figure 14. This finding suggests that as warning reliability increased, then participants became more sensitive and better able to distinguish between warnings that functioned properly and those that malfunctioned. Paired sample *t*-tests conducted to evaluate differences between PPV condition means found significant differences among all three PPV levels at the $p < 0.05$ level (PPV = .40: *M* = 2.16, *SD* = 1.19; PPV = .60: *M* = 2.91, *SD* = .84; PPV = .83: *M* = 3.47, *SD* = .58).

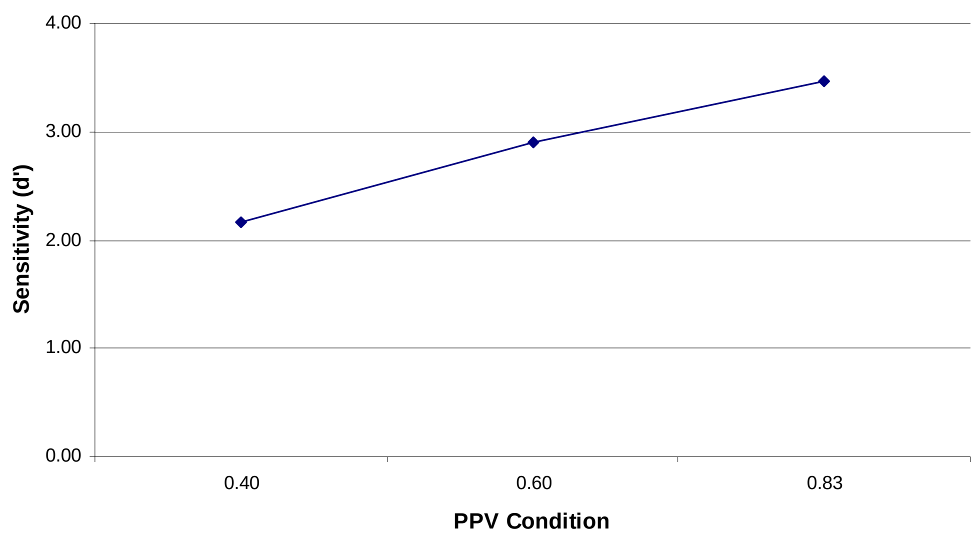


Figure 14. Mean Sensitivity as a Function of Warning System Reliability (PPV) in the Priming Task.

Another way to view sensitivity differences across PPV conditions is to graph an ROC curve that depicts both participants' hit rate (i.e., the valid stop rate) and the false alarm rate (i.e., the false stop rate). Figure 15 presents an ROC plot derived from the priming task (note that the ROC plot also depicts points from the driving task; an interpretation of these data points and a comparison of the data points from the priming and driving tasks will be discussed in more detail in Section 3.2.4.3). In the figure, the hit rate is plotted as a function of the false alarm rate for each PPV level. As noted previously, the diagonal line spanning the entire chart represents chance performance. Sensitivity improves as performance moves from the center of the chart to the upper left corner (along the minor diagonal). The figure shows that as sensitivity increased in the priming task (the blue points), the curves shift toward the upper left corner. The pattern shown in the ROC curve is consistent with the results from the ANOVA analyses indicating that participants became more sensitive as PPV increased.

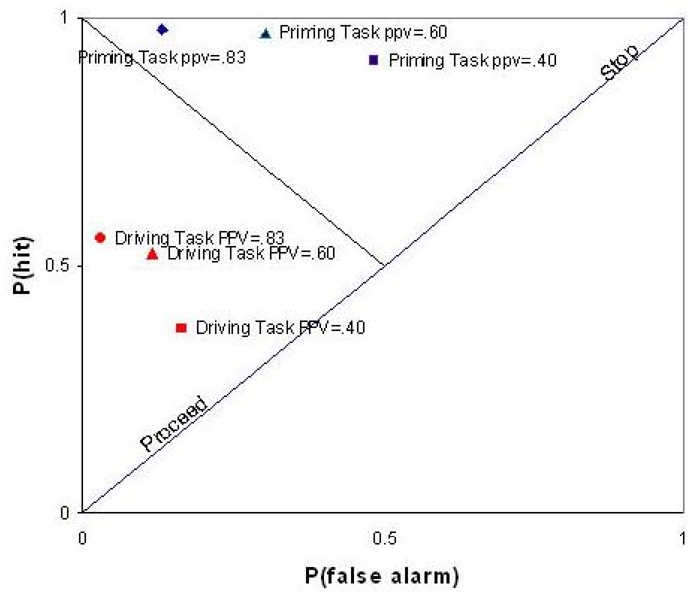


Figure 15. ROC Points Obtained in the Priming Task.

3.2.2.3 Decision Criterion (Response Bias) in the Priming Task

Response bias captures a tendency either to comply with the warning, or to violate it. Changes in response bias can be seen by the pattern of data points from the ROC curve in Figure 15; performance to the right of the minor diagonal reflects a bias to stop, while performance to the left reflects a bias to proceed through the crossing. The relative position of the three blue points in the ROC curve shown in Figure 15 suggests a bias towards stopping at grade crossings. An examination of the changes in response bias across PPV conditions, as illustrated in Figure 16, report a similar effect; as warning reliability decreased, participants' became more conservative (i.e., more likely to stop), $F(2,48) = 6.76$, $p < .01$. Paired samples *t*-tests comparing PPV condition means found a significant difference between all three rates at the $p < 0.05$ level (PPV = 0.40: *M* = -0.88, SD = 0.14; PPV = .60: *M* = -.77, SD = .47; PPV = .83: *M* = -.55, *SD* = .28).

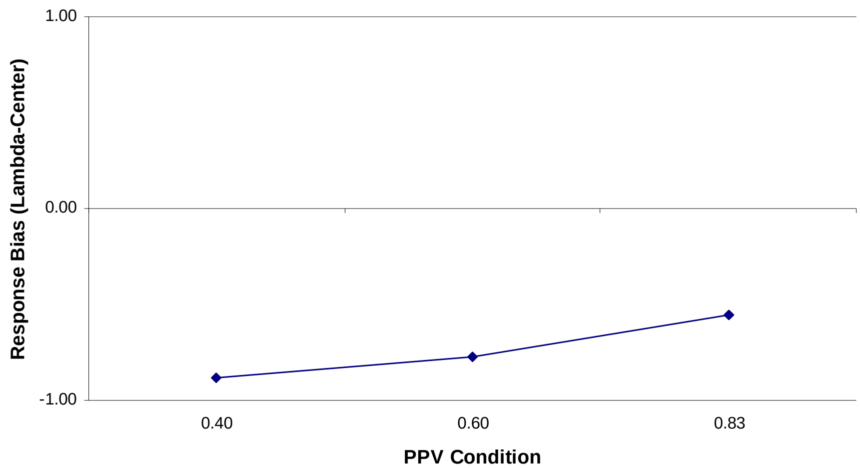


Figure 16. Participants' Mean Response Bias as a Function of PPV in the Priming Task.

3.2.3 Experiment 1 and Experiment 2-Priming Task Comparison

In Experiment 1 and in the priming task used in Experiment 2, participants responded to static images of active grade crossings by indicating whether they would proceed or stop. The two tasks were similar in that participants were presented with very ambiguous situations and asked to make quick judgments. However, the two tasks differed in three ways. First, in Experiment 1, the PPV rate was manipulated by presenting false warning activations only, whereas in Experiment 2, the PPV rate was manipulated by presenting not only false warning activations but also warning misses (i.e., when the warning failed to indicate a train's arrival). Second, the sound of a train horn was provided in Experiment 2 as a secondary cue to indicate train arrival. Third, performance feedback, which was provided in Experiment 1, was not presented in Experiment 2.

Despite these differences, it was of interest to compare participants' performance in Experiment 1 to that in the priming task in Experiment 2 to examine how these changes influenced participants' sensitivity and response bias.

3.2.3.1 Sensitivity

Changes in sensitivity between Experiment 1 and the priming task in Experiment 2 can be seen by comparing the pattern of points in the ROC curves shown in Figure 5 (Experiment 1) with that in Figure 15 (Experiment 2). Figure 17 below summarizes the differences in sensitivity for the high, medium, and low PPV rates. Because Experiment 1 did not have a 0.83 PPV rate condition, the data shown in the "high" PPV rate category is for the 0.87 PPV rate.

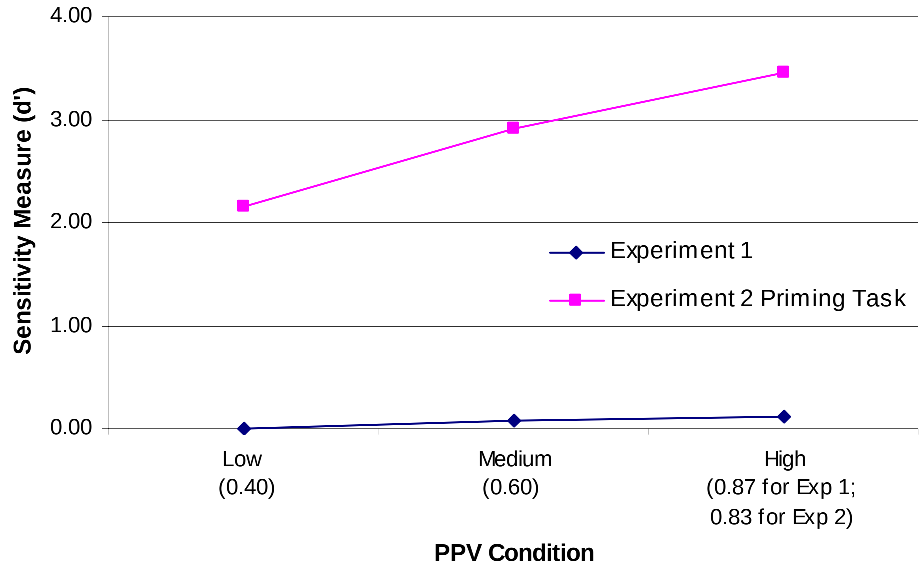


Figure 17. Differences in Sensitivity Obtained in Experiment 1 and in the Priming Task in Experiment 2

As Figure 17 shows, participants in Experiment 2 were better able to differentiate reliable from unreliable warnings than those in Experiment 1. One factor may be the availability of a train horn, which provided an additional cue that train arrival was imminent. Although the "signal" for the task was operationally defined as a lowered gate, the lack of performance feedback may have led participants to choose for themselves what they considered the signal to be. As a result, they could rely on either the visual cue of the lowered gate, which varied in its reliability, or the auditory cue of a train horn, which was always reliable, and choose their responses accordingly.

To test this hypothesis, we calculated participants' sensitivity on those trials in which the horn was sounded in each of the three PPV conditions, and compared those values to the sensitivity on all other trials when only the gate was shown. Figure 18 shows changes in sensitivity as a function of warning system reliability (the PPV rate) and "signal" (i.e., the gate versus the sound of the train horn). Note that in Figure 18, the data for the "horn" trials reflect the condition in which it was used, and not the PPV rate of the horn. The sound of the horn was always reliable.

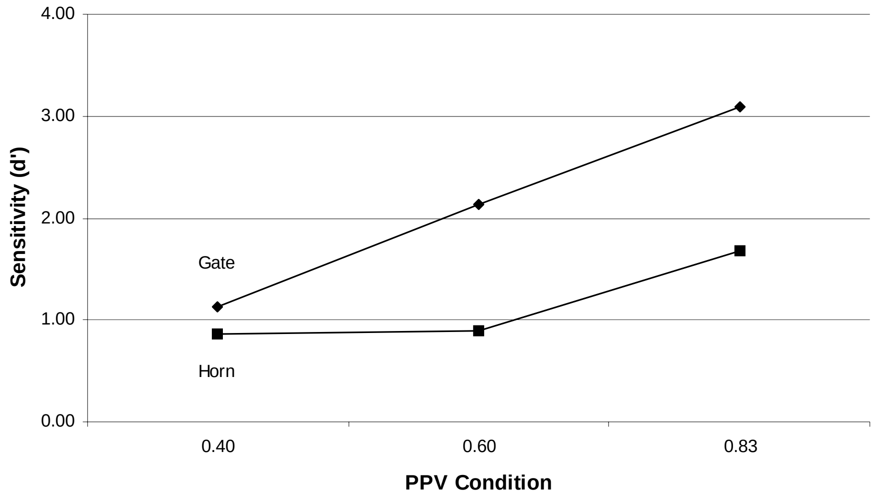


Figure 18. Changes in Sensitivity as a Function of PPV for Signal Defined as Gate and Horn

Changes in participants' sensitivity were influenced more by the PPV rate when participants used the visual cue of the lowered gate as the signal rather than the sound of the horn. When participants relied on the train horn as their signal, their sensitivity remained unchanged between PPV = .40 and PPV = .60, but rose dramatically from PPV = .60 to PPV = .83. The results suggest that when participants defined their "signal" consistent with that specified by traffic law, participants' sensitivity increased as warning system reliability increased, as expected. The sound of the train horn was helpful only when warning system reliability was high.

3.2.3.2 Response Bias

A comparison of the ROC curves in Figure 5 (Experiment 1) and Figure 15 (Experiment 2) shows that participants in Experiment 2 were generally more conservative than those in Experiment 1, and were more likely to stop at the crossing. Similar to the sensitivity data, it was of interest to examine how the sound of the train horn contributed to this change in responding behavior. We calculated the response bias on those trials in which the horn was sounded in each of the three PPV conditions and compared them to the response bias for those trials in which only the gate was shown. Figure 19 shows changes in response bias as a function of warning system reliability (PPV) and signal (i.e., the gate versus the sound of the train horn). Note that in the figure, the data for the horn trials reflect the condition in which it was used, and not the PPV rate of the horn, because the sound of the horn was always reliable

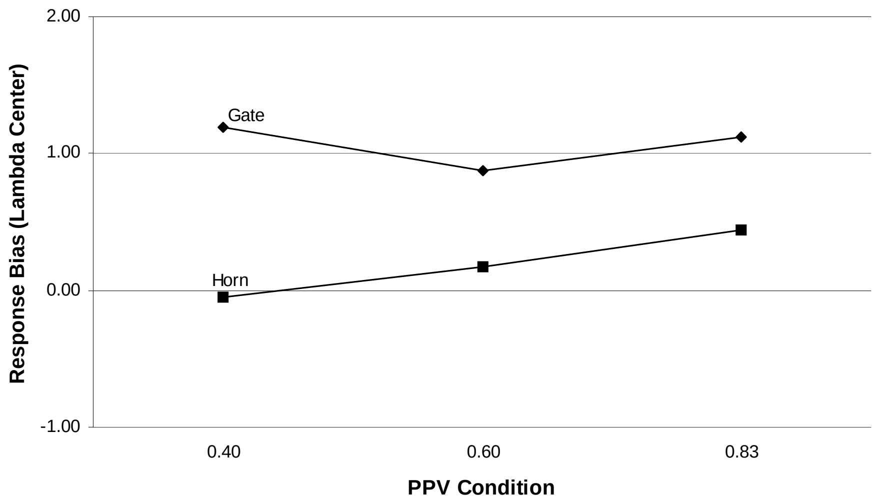


Figure 19. Changes in Responses Bias as a Function of PPV for Signal Defined as Gate and Horn

The figure shows that participants who used the gate as their signal were generally more likely to stop at the crossing than those who used the train horn as their signal. The figure also suggests that participants who used the train horn as their signal modified their response criterion as a function of the PPV rate. Therefore, they became less conservative, and less likely to stop at the crossing, as the PPV rate decreased if they did not hear the sound of the train horn. This behavior is consistent with expectations; as participants perceived the active warning system becoming less reliable, they relied on what they perceived to be the more accurate cue, the train horn. However, changes in participants' response criterion, when they used the lowered gate as the signal, are less clear. The data in Figure 19 indicates that participants were most conservative—that is, most likely to stop—when the PPV rate was at its highest (0.83) and at its lowest (0.40).

3.2.4 Driving Task Performance

Similar to the data collected in the priming task, driving task behavior was measured via compliance (PCR), sensitivity (d′), and response bias (λ_c). Details for each participant's performance (i.e., their rate of hits, misses, false alarms, and correct rejections) are provided in Appendix E. In addition to these three variables, driving performance was evaluated in terms of frequency of collisions with other objects (e.g., trains and gates), driving task completion time, train time to crossing, and frequency of gate violations. These four measures of driving performance are independent of the sensitivity and response bias measures and provide supplementary information about participants' behavior when presented with ambiguous warning signals.

In this section, we first consider the effects of warning reliability on driving behavior in the simulator by examining changes in participants' compliance, sensitivity, and response bias as a function of PPV rate. We compare participants' behavior, defined by these three measures, in the driving task to that in the priming task. We then examine the effects of warning reliability on driving performance.

3.2.4.1 Compliance in the Driving Task

Compliance, defined by PCR, was assessed separately for each warning system state based on the extended SDT matrix shown in Table 6. Although this approach is somewhat unconventional, it allows for an examination of changes in compliance with warnings that functioned properly and with warnings that malfunctioned. Previous research has shown that compliance with a warning system that was properly activated fostered warning trust, whereas compliance with false warning activations led to warning mistrust, and that this mistrust in the warning system could contribute to disregarding future warnings (e.g., Chugh and Caird, 1999; Parasuraman and Riley, 1997).

A 3 (PPV rate: .40, .60, .83) x 2 (warning state: proper function vs. system failure) ANOVA was performed on the PCR data to investigate if a significant difference existed between warning system states. Figure 20 shows compliance as a function of the PPV rate when the warning system functioned properly and when it failed.

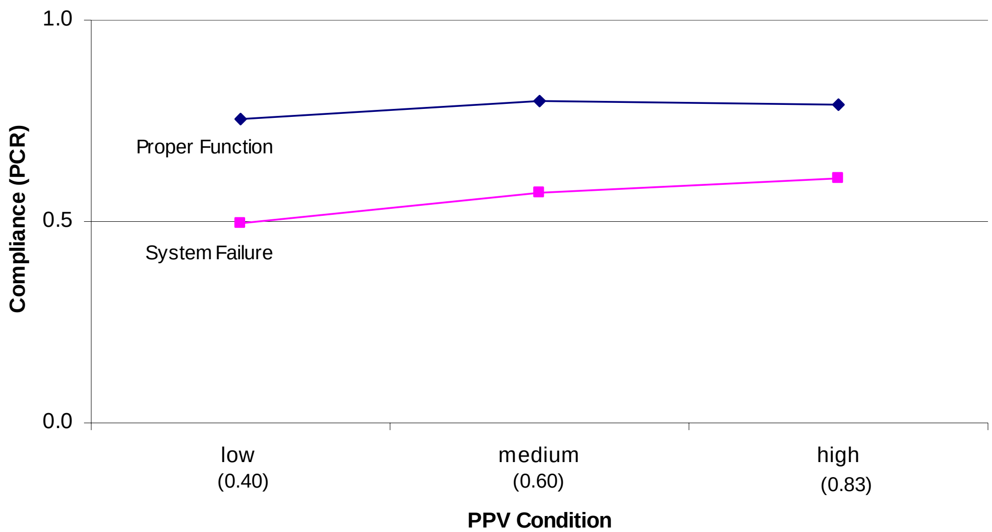


Figure 20. Mean Warning Compliance (PCR) as a Function of Warning System Reliability (PPV) and Warning System State in the Driving Task
The terms "low," "medium," and "high" refer to participants' expectations created in the priming task and correspond to the three PPV conditions.

As expected, the results showed that compliance significantly differed as a function of warning system state, $F(1,144) = 48.84$, $p < .001$. We also expected that compliance in the driving task would increase as PPV increases in both warning system states, but there was no overall effect of PPV rate on PCR, $F(2,144) = 2.02$, $p > .05$, nor was there a significant interaction between warning system state and PPV, $F(2,144) = .47$, $p > .05$. Although the data shown in Figure 20 indicates that compliance to the warning signals was high regardless of PPV rate when the system functioned properly, there is strong visual evidence to suggest that PPV rate affected compliance when the warning system failed, despite the lack of a significant interaction. As a result, a repeated measure ANOVA was conducted on the data for warning system failure only with PPV rate as the dependent variable. This analysis found a significant PPV effect on compliance (PCR), F (2, 48) = 3.48, $p < .05$; paired-sample *t*-tests using the Bonferroni adjustment showed that compliance was greater at high PPV rates (PPV = .83; *M* = 0.61, *SD* = .19) than at low PPV rates (PPV = .40; *M* = .50, *SD* = .15, *t* (24) = -2.91, $p < .02$).

3.2.4.2 Sensitivity in the Driving Task

A within-subjects repeated measures ANOVA examining changes in participants' sensitivity (d′) as a function of the PPV rate revealed that participants' ability to distinguish between warnings that functioned properly and those that malfunctioned improved as warning reliability increased, $F(2,48) = 13.72$, $p < .001$. Figure 21 illustrates changes in sensitivity as a function of the overall PPV rate; the actual reliability of the warning signal was not considered in this analysis. Paired sample *t*-tests conducted to evaluate differences between PPV condition means found a significant difference between all three PPV rates at the p < 0.05 level (PPV = .40: *M* = 1.12, *SD* = 1.19; PPV = .60: *M* = 2.13, *SD* = 1.56; PPV = 0.83: *M* = 3.09, SD = 1.88).

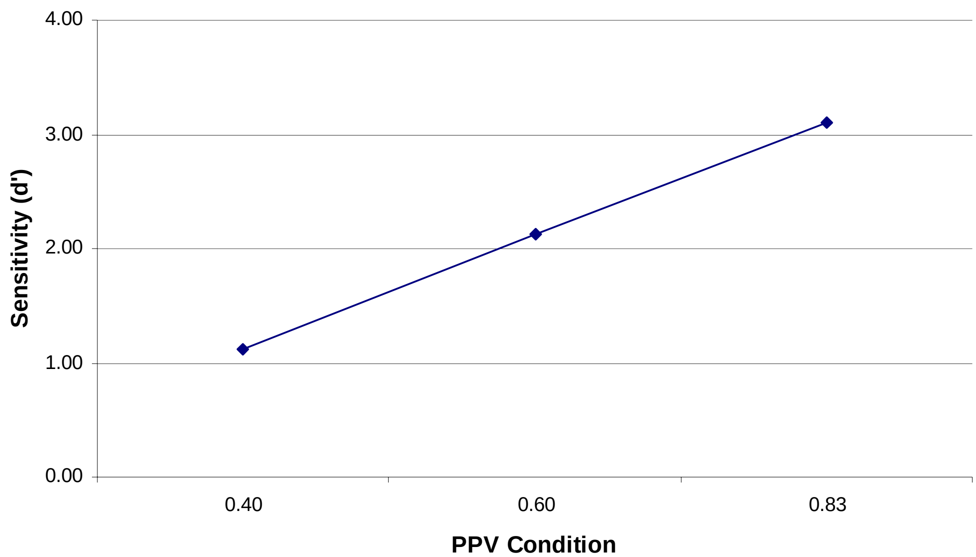


Figure 21. Mean sensitivity as a function of warning system reliability (PPV) in the driving task.

Changes in sensitivity can be better seen in the ROC curve, shown in Figure 15. Three points for the driving task, one for each PPV rate, are shown in red. As the figure shows, participants became better at distinguishing reliable from unreliable warnings as the PPV rate increased; participants were most sensitive when the PPV rate was high (.83) and least sensitive when it was low (.40).

3.2.4.3 Decision Criterion in the Driving Task

We expected that participants would become less conservative in the driving task as PPV decreased. However, contrary to our expectations, no significant effect of PPV rate was found. In fact, the relative position of all three points for the driving task in the ROC curve in Figure 15 suggests a bias to proceed through the grade crossing. Interestingly, a comparison of driving behavior between the priming task and the driving task highlights participants' propensity to proceed through the crossing in the driving task relative to the priming task. This result may be attributable to the financial incentive participants were offered for completing the driving course quickly.

3.2.4.4 Descriptive Driving Performance Measures

Data for the frequency of collisions (with trains and gates), driving task completion time, train time to crossing, and frequency of gate violations were evaluated to provide a better understanding of motorist behavior in the simulator. These measures are independent of the compliance, sensitivity, and response bias variables described above. Instead, the purpose of examining these driving performance measures was to learn what behaviors participants engaged in when they were presented with ambiguous warning signals.

Frequency of Collisions

Table 7 shows the number of vehicle-train and vehicle-gate collisions for each PPV rate.

Table 7. Driving Task Collision Types and Frequencies

	Condition		
	PPV = .40	**PPV = .60**	**PPV = .83**
Vehicle-Train Collision	2	0	1
Vehicle-Gate Collision (Reliable Warning Signal)	0	10	4

In examining the table, we first consider the frequency of vehicle-train collisions (the first row in Table 7). When the PPV rate was .40, 2 vehicle-train collisions occurred. In one instance, the collision was the result of a high-risk violation; the warning system was activated and accurate, but the participant violated the lowered gates, possibly due to a perception of low signal credibility due to the low PPV rate of the warning signals experienced at previous crossings. The other collision resulted because the warning system failed to provide reliable information about an imminent train arrival (i.e., a warning system miss). No vehicle-train collisions occurred in the PPV = .60 condition, and only one collision occurred when PPV = .83.

Vehicle collisions with lowered gates were found to be more prevalent than vehicle-train collisions. Ten of these collisions were observed in the PPV = .60 condition; eight of the collisions in this condition resulted from high-risk violations to reliable warning signals. Four vehicle-gate collisions occurred when the PPV rate was 0.83. Of these collisions, three were due to high-risk violations to reliable warning signals. Some vehicle-gate collisions were due to excessive speed and some were the result of poor vehicle maneuvering around lowered gates. Interestingly, no vehicle-gate collisions occurred when the PPV rate was lowest (.40).

Task Completion Time

A comparison of the time participants took to complete the driving task as a function of PPV rate showed no difference across the three conditions ($p > 0.05$). On average, the participants drove the course in 22 minutes. There were a few exceptions, however; two participants drove through the driving course without complying with any of the warning signals to maximize their payoff benefit and minimize their time commitment.

In a sense, the data reveal that stopping at a crossing, as participants were more likely to do when the PPV rate was high, did not significantly increase the time required to reach the destination. That is, there was no cost to compliance. However, there were costs to collisions in the simulator; in addition to a crash delay of about 5 seconds, participants incurred another 5-second delay associated with resuming the simulation and getting the vehicle back on the course.

Train Time to Crossing

Train time to crossing measured the time in seconds that it took for a train to reach a grade crossing after the vehicle crossed the tracks. This measure assessed the relative risk of driving through the grade crossing in front of an oncoming train. On average, participants violated the gates 4.28 seconds before the train's arrival across all PPV conditions. There was no difference in the mean train time to crossing as a function of PPV rate.

Frequency of Gate Violations

We examined the frequency of gate violations based on whether the warning system was reliable or not. When the warning system accurately signaled the approach of a train, participants committed violations 71 percent of the time when the PPV rate was high (0.83), 71 percent of the time when the PPV rate was at a medium level of reliability (0.60), and 74 percent of the time when the PPV rate was low (0.40) (this difference among PPV levels was not significant at the $p < 0.05$ level). We then classified the violations into two types:

- Type A violations: Motorists stopped the vehicle at the lowered gate and waited, but then proceeded through the crossing and violated the gates prior to the train's arrival.
- Type B violations: Motorists stopped the vehicle at the lowered gate, waited for the train to pass through the grade crossing, but then violated the lowered gate before it was fully raised. By doing so, motorists disregarded the possibility that multiple tracks may be present at the grade crossing and that a second train may arrive.

Both Type A and Type B violations are dangerous, and some probability of an accident is associated with each. There was no difference in the frequency of Type A or Type B violations as a function of the PPV rate, as shown in Table 8.

Table 8. Mean Proportion of Type A and Type B Violations as a Function of Warning System Reliability

	Condition		
	PPV = .40	PPV = .60	PPV = .83
Compliant (valid stop)	26%	29%	29%
Type A	37%	29%	36%
Type B	38%	42%	34%

When the warning system was unreliable, participants committed violations 91 percent of the time when the PPV rate was low, 80 percent of the time when the PPV rate was at its medium level, and 66 percent of the time when the PPV rate was high. A within-subjects ANOVA indicated an overall effect of PPV rate on compliance, $F(2, 74) = 3.2$, $p < 0.05$; post-hoc comparisons showed that there were significantly more violations committed when the PPV rate was 0.40 than when it was at its highest at 0.83 ($p < 0.05$).

To understand motorist behavior when the warning system was unreliable, we examined when the violations occurred by redefining the Type A and Type B violations to take into account that no train was arriving.

- Type A violations: Motorists failed to stop when the warning signal was activated and intentionally violated the lowered gates.
- Type B violations: Motorists stopped the vehicle but then violated the lowered gates before they were raised.

Table 9 shows the proportion of Type A and Type B violations. The data suggests a trend for the proportion of Type A violations (i.e., proceeding without stopping) to decrease as the PPV level increased, but these differences were not significant at the $p < 0.05$ level. Similarly, there was no difference in the proportion of Type B violations across PPV conditions.

Table 9. Mean Proportion of Type A and Type B Violations as a Function of Warning System Reliability

	Condition		
	PPV = .40	**PPV = .60**	**PPV = .83**
Compliant (false stop)	9%	20%	34%
Type A	73%	58%	48%
Type B	19%	22%	18%

3.2.5 Summary of Findings

The purpose of Experiment 2 was to assess compliance with active warning devices as a function of their reliability in a dynamic environment. Participants were primed to the reliability of the warning system in a task similar to that used in Experiment 1. The PPV rate was set at three levels (0.40, 0.60, and 0.83), which was manipulated by the inclusion of warning system false alarms and warning system misses. Examination of the data from the priming task showed that compliance increased as PPV rate increased when the warning system was reliable. Additionally, participants became more sensitive to the accuracy of the warning system as reliability improved, but they were more conservative in their responses when the reliability of the warning system was low. The results of the priming task are consistent with previous research showing that operators matched their likelihood of responding to the reliability of the warning system (Bliss et al. 1995; Getty et al., 1995; Maltz and Meyer, 2001; Meyer, 2001) and demonstrate the detrimental effects of reduced warning reliability on warning response performance (i.e., noncompliance).

The consequences of warning unreliability were further demonstrated in the driving task, where participants were more compliant to reliable warnings than unreliable ones. There was no overall effect of changes in the PPV rate on the proportion of participants' correct responses, but evidence suggested that decreasing the PPV rate reduced compliance to unreliable warnings. Similar to the results of the priming task, an examination of participants' sensitivity in the driving task revealed that as the PPV rate increased, participants' sensitivity to reliable versus unreliable warnings also improved. However, changes in the PPV rate did not significantly change participants' response bias; participants were inclined to proceed (a liberal criterion with positive λ_c values), regardless of the warning reliability. This finding is different from that reported by Raslear (1995), who suggested that the presence of active warning devices at grade crossings biases motorists to stop and comply when activated. The difference may be attributable to the incentives offered to participants to complete the driving task as quickly as possible, which may have encouraged them to commit the violations and offers insight into how participants' motivations dictate behavior.

Descriptive driving performance measures (i.e., collision frequency, driving task completion time, train time to crossing, and frequency of gate violations) provided additional information about motorists' behavior. Of these measures, data for the frequency of gate violations provided the most compelling evidence of the costs of unreliable warnings; as the PPV rate decreased, the frequency of gate violations increased. Interestingly, a comparison of driving task completion time indicated that violating the gates did not significantly reduce the time required to reach the destination, despite the fact that a few participants drove through the course without complying with any of the warning signals.

The results of Experiment 2 support the hypotheses set forth in Lerner, et al. (1990) and that is explored in more detail in Yeh and Multer (in preparation) that warning system unreliability could have a detrimental and predictable effect on motorists. As motorists' perceive the warning system to be less credible, they will be more likely to violate the warning signal, perceiving little risk to their safety because the warning system has failed before.

4. General Discussion

The purpose of the two experiments was to examine shifts in participants' response strategies at active grade crossing warnings as a function of varying reliability levels. In Experiment 1, we used a probability learning approach to assess whether participants were sensitive to reduced warning reliability. We examined participants' responses to eight PPV rates, which were manipulated through the inclusion of warning false alarms. The results showed that as the PPV rate decreased, participants' compliance with the warning signal decreased as well. Participants matched their responses to the predictive ability of the warning system, which was consistent with that reported in previous research by Bliss, et al. (1995). However, because participants were compliant with reliable warning systems, when those systems failed, participants were likely to respond with a false stop. This behavior contributes to motorist distrust in the warning system (i.e., the "cry-wolf" effect) and can lead to future noncompliance (Breznitz, 1983; Bliss, et al., 1995; Pate-Cornell, 1986).

We were surprised that participants in Experiment 1 were not sensitive to changes in the PPV rate, unless the PPV rate dropped from near-perfect reliability (from a PPV rate of 0.97 to 0.87). However, it is possible that the signal was not salient enough; in the study, the presence of a signal needed to be inferred from knowledge of the PPV rate and feedback from prior trials. Thus, the results suggest that if motorists perceive that a warning system is unreliable, they may violate the warning system, if they believe the gates were lowered in error.

In Experiment 2, we used a signal detection paradigm to evaluate motorist compliance with actively protected grade crossings in a simulated driving environment. In contrast to Experiment 1, the PPV rate was manipulated by the inclusion of warning false alarms and warning system misses. Participants completed a priming task to prime them to the PPV rate of the warning system before driving a vehicle through a course in a simulator. The results of the priming task showed that similar to Experiment 1, compliance decreased as PPV decreased when the warning system was unreliable, and that participants were generally conservative in their responses (i.e., more likely to stop) as warning reliability decreased. Unlike Experiment 1, however, participants in Experiment 2 were sensitive to changes in the PPV rate as warning reliability improved; that is, participants were able to distinguish reliable from unreliable warnings.

One change to the methodology used in Experiment 1 for the priming task in Experiment 2, which could account for this difference in sensitivity, was the addition of the sound of a train horn to indicate imminent train arrival. The train horn provided an auditory cue in conjunction with the visual cue of the lowered gate, but unlike the visual warning, the auditory warning was perfectly reliable. Although we defined the "signal" to be the visual cue of the lowered gate to be consistent with traffic law, the presentation of two warning cues in the second experiment allowed participants to define for themselves what they considered the signal to be. When participants considered the lowered gate to be their signal, their sensitivity increased as the PPV rate increased, as expected. However, when participants relied on the train horn, there was no change in sensitivity when the PPV rate was low (e.g., when the PPV rate decreased from 0.60 to 0.40), but the auditory cue was helpful when the reliability of the warning system was high (i.e., when the PPV rate increased from 0.60 to 0.83).

Participants' compliance to warnings as a function of their reliability was further examined in Experiment 2 using a driving task in which participants maneuvered a vehicle through a simulated course and decided whether to stop or proceed when they encountered grade crossings. Participants were more compliant at crossings when they perceived the warning system to be reliable than when they perceived it to be unreliable. Similar to the priming task, participants were sensitive to changes in the PPV rate, such that lowered reliability resulted in reduced compliance. However, participants did not adjust their response criterion with respect to the PPV rate but were inclined to proceed regardless of the reliability. Descriptive data measuring driving performance showed that the frequency of gate violations increased as the PPV rate decreased. Unfortunately, the manipulation of warning system unreliability in Experiment 2 was such that we were not able to directly compare the effects of warning signal misses and false alarms on motorist behavior as done in several studies described in the literature review (Chugh and Caird, 1999; Cotté, et al., 2001; Dixon, et al., 2004; Maltz and Meyer, 2001; Maltz and Shinar, 2004; Meyer, 2001).

The results of the two experiments suggest that motorists are sensitive to reliability or credibility of warning information provided at actively protected grade crossings and highlight the importance of high warning signal credibility in encouraging compliance at grade crossings. The results are consistent with that reported in the theoretical literature (e.g., Bliss, et al., 1995; Getty, et al., 1995; Maltz and Meyer, 2001; Meyer, 2001) and show a pattern of behavior similar to that reported in field studies of motorist behavior at active grade crossings (Wilde, et al., 1987).

5. Recommendations and Future Research

The results of the present studies suggest that improving motorists' perception of signal reliability may reduce gate violations. Stopping at a grade crossing has "costs" in terms of delays, and stopping unnecessarily leads to motorist frustration. In the experiments conducted here, signal reliability was achieved by reducing the rate of false alarms (Experiments 1 and 2) and missed signals (Experiment 2).

From an engineering perspective, the false alarm rate can be reduced through improvements in track circuitry and train detection equipment, incorporating good maintenance practices, and identifying and correcting signal malfunctions in a timely manner (FHWA, 2002). From a cognitive science perspective, additional research is needed to investigate factors that motorists use to assess warning system credibility. We recommend research that examines the value provided by external sources of information regarding a train's arrival at a grade crossing. In addition to an activated warning device, a train's arrival may be indicated visually by the presence of the train or the beam of its lights, or it may be indicated aurally by the sound of a train horn or wayside horn. These external cues may help motorists discriminate between reliable and unreliable warnings. The validity of these cues varies, and it is of interest to determine which of these cues motorists' judge to be most reliable. For example, some participants in Experiment 2 relied on the sound of the train horn rather than the visual cue of the lowered gate to make their decision to stop or proceed. Although the auditory warning was effective in the experiment conducted here, additional research should examine motorists' use of auditory cues when the warning comes from a wayside horn and thus less reliable than the train horn.

Additionally, motorists' perception of the value of warning information can be observed in their responses to the warning signal (i.e., whether or not they comply), but it can also be calculated as a function of the difference between the expected value of information when a warning is presented and the expected value of information when no warning is presented. If the expected value of the warning information is greater than the cost of looking for other cues regarding an imminent train arrival, then the motorist will be more likely to comply. Decisionmaking models can be applied to better understand how motorists' calculate these values. In the warning literature, examination of operators' decisions to comply with warnings identified two models that may be applicable. One model is based on the hypothesis that motorists use a Bayesian approach to decisionmaking and quickly calculate the likelihood of an oncoming train from the expected value of warning cues and visual cues. Another model is based on the hypothesis that motorists use a "take the best" decision heuristic in considering and comparing the reliability of the cues that are available and decide whether to stop or proceed from what they perceive to be the most reliable cue (Lehto, 2006). The results of these research efforts would benefit warning designers by providing recommendations on how to set the sensitivity for a warning system.

Another area for research is to examine the interaction between the motorist and warning signal using a model of distributed team signal detection. Lehto (2006) proposed that the operator and the warning system can be considered to be a team working to reach a joint decision to optimize performance. The distributed team signal detection model predicts that changes by one "team member" will alter the response criterion for the other to optimize performance. The human operator's decision making, as proposed by the distributed team signal detection model, builds on the fact that the human operator combines various sources of information to assess the warning signal's reliability. The results of Experiment 2 provide evidence of this behavior: participants used the sound of a train horn or visual information regarding whether a train was approaching the crossing to optimize their behavior. With this information, participants were more sensitive to the PPV rate of the warning signal and were more likely to comply when they perceived the signal to be reliable.

The distributed team signal detection model suggests that warning designers could use motorists' behavior to determine how stringently to set the decision threshold for a warning system. Observing the overall rate of compliance at crossings would provide information regarding whether motorists find the signal to be reliable or not and would serve as an indication as to whether a shift in the warning signal's operating criterion is required. In particular, observing drivers, who frequent the crossing daily and are familiar with it, would be valuable, because this population has developed expectations regarding the likelihood of a train's arrival at the crossing. Drivers who are unfamiliar with the crossing are generally more cautious than drivers familiar with the area and are less likely to commit a violation (Lerner, et al., 1990). Thus, research that empirically examines motorists' performance with respect to the warning system using this team framework and that explores how the human operator and warning system can cooperatively optimize team performance would be valuable.

It is also worthwhile to obtain a better understanding of motorists' cost-benefit structures on their response at a crossing. Although a weighting of costs and benefits on operator warning response strategies has been investigated in more general situations (e.g., Bliss et al., 1995; Getty et al., 1995; Edworthy, 2000; Edworthy & Dale, 2000; Mellor, Holzworth & Conway, 2003), it is not clear how motorists evaluate the costs and benefits associated with compliance behavior at active grade crossings. Future research should focus on multiple cost-benefit structures associated with compliance with active grade crossing warning devices.

Finally, we recommend an investigation of the degree to which motorists' expectations regarding the likelihood of a train at a crossing play a role in compliance. Familiarity with the particular crossing has been related to both dangerous actions and actual accident involvement (Lerner et al. 1990). For example, motorists who do not encounter trains as they drive to work during their regular commuting times may not expect to encounter trains at these grade crossings when traveling at other times of the day. The probability of motorist noncompliance and accidents increases when these expectations are violated. Thus, future research should focus on sequential dependencies to investigate the degree to which violation of one's expectations about train arrival impact compliance.

APPENDIX A: Proportions of Valid stops and false stops obtained in Experiment 1

Noncompliance (High and No-Risk Gate Violations)

Two response categories (valid stops and false stops) provide independent information about participant's responses (Macmillan and Creelman, 2005). Since the proportion of valid stops is already known, high-risk violation proportion can be calculated by subtracting the valid stop proportion from 1. Investigating high-risk violations is an alternative way of examining valid stops, if one is interested in noncompliant rather than compliant responses. Similarly, low-risk violation proportions can be calculated by subtracting false stop proportion from 1, and proportion of overall gate violations can be calculated by subtracting overall compliance proportion from 1. Illustrations of these responses are presented next. These measures were not subjected to significance testing because they do not provide independent information from what has been already reported, however, they are useful in understanding changes in responses from an alternative perspective.

The tables below show the valid stop and false stop proportions obtained in Experiment 1.

Table A-1. Valid stop proportions obtained in the static task (N = 8).

Valid Stop Proportions

PPV Rate	s-1	s-2	s-3	s-4	s-5	s-6	s-7	s-8	Mean	SD
0.23	0.0714	0.6000	0.4714	0.2571	0.1000	0.7857	0.1571	0.2000	0.3304	0.2598
0.30	0.2556	0.7111	0.9333	0.4000	0.4444	0.5667	0.6667	0.4778	0.5569	0.2111
0.40	0.3333	0.6500	0.9000	0.3667	0.4667	0.9917	0.3583	0.3833	0.5563	0.2617
0.60	0.7833	0.7722	0.9944	0.9944	0.5444	0.9944	0.6000	0.6111	0.7868	0.1905
0.70	0.8333	0.7381	0.9952	0.9952	0.6000	0.9952	0.8952	0.7667	0.8524	0.1452
0.77	0.8652	0.7609	0.9261	0.9957	0.9261	0.9783	0.9696	0.7391	0.8951	0.0984
0.87	0.9423	0.8077	0.9962	0.9962	0.9500	0.9500	0.9577	0.8500	0.9313	0.0674
0.97	0.9931	0.9828	0.9690	0.9966	0.9448	0.9966	0.9621	0.9862	0.9789	0.0187

Table A-2. False stop proportions obtained in the static task (N = 8).

False Stop Proportions

PPV Rate	s-1	s-2	s-3	s-4	s-5	s-6	s-7	s-8	Mean	SD
0.23	0.1087	0.5739	0.5043	0.2478	0.1478	0.8087	0.1783	0.2261	0.3495	0.2504
0.30	0.3381	0.6571	0.8952	0.3619	0.4429	0.5667	0.6571	0.3571	0.5345	0.1962
0.40	0.3444	0.6111	0.8611	0.4389	0.4333	0.9944	0.3389	0.4000	0.5528	0.2488
0.60	0.7583	0.8250	0.9917	0.9750	0.6583	0.9917	0.6167	0.5083	0.7906	0.1869
0.70	0.7778	0.7667	0.9889	0.9889	0.5667	0.9889	0.8889	0.8000	0.8458	0.1484
0.77	0.8857	0.6857	0.9143	0.9857	0.8714	0.9857	0.9857	0.7429	0.8821	0.1142
0.87	0.9750	0.9000	0.9750	0.9750	0.9000	0.9750	0.9000	0.8750	0.9344	0.0442
0.97	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.9000	0.0000

Figure A-1 demonstrates that the proportion of high-risk violations decreased as a function of warning system reliability (PPV). Proportions of high-risk violations were not subjected to significance testing because they are dependent on valid stop proportion. However, because a significant PPV effect was found on valid stop proportions (see Figure 7), it is reasonable to assume that an analogous analysis of high-risk violations would produce significant results.

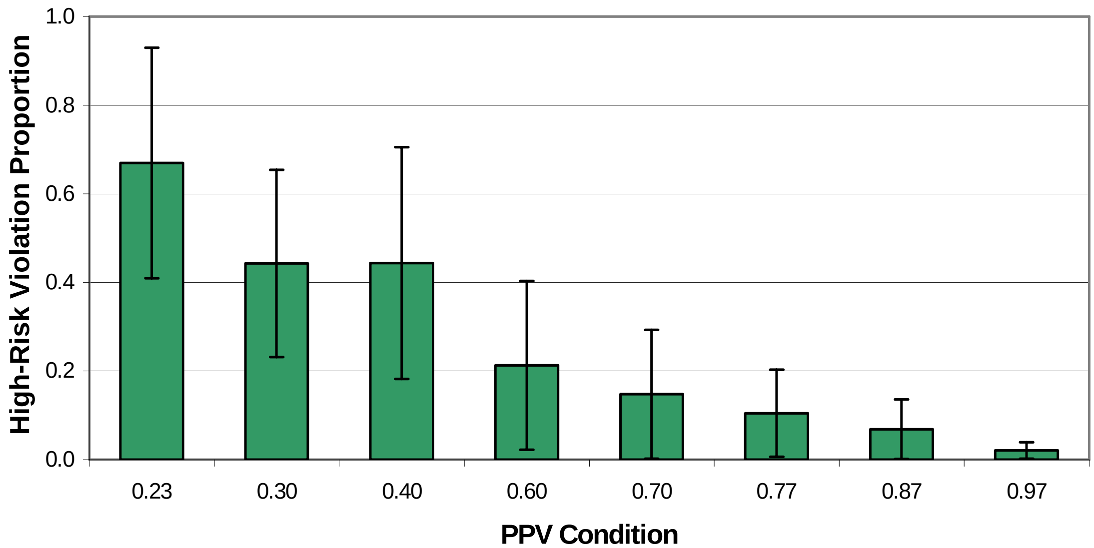


Figure A-1. Means and Standard Deviations for High-Risk Violation Proportions

Figure A-2 demonstrates that proportion of no-risk violations decreased as a function of warning system reliability (PPV). Proportions of no-risk violations were not subjected to significance testing because they are dependent on false stop proportion. However, since a significant PPV effect on false stop proportions was found (see Figure 9), it is reasonable to assume that an analogous analysis of no-risk violations would produce significant results.

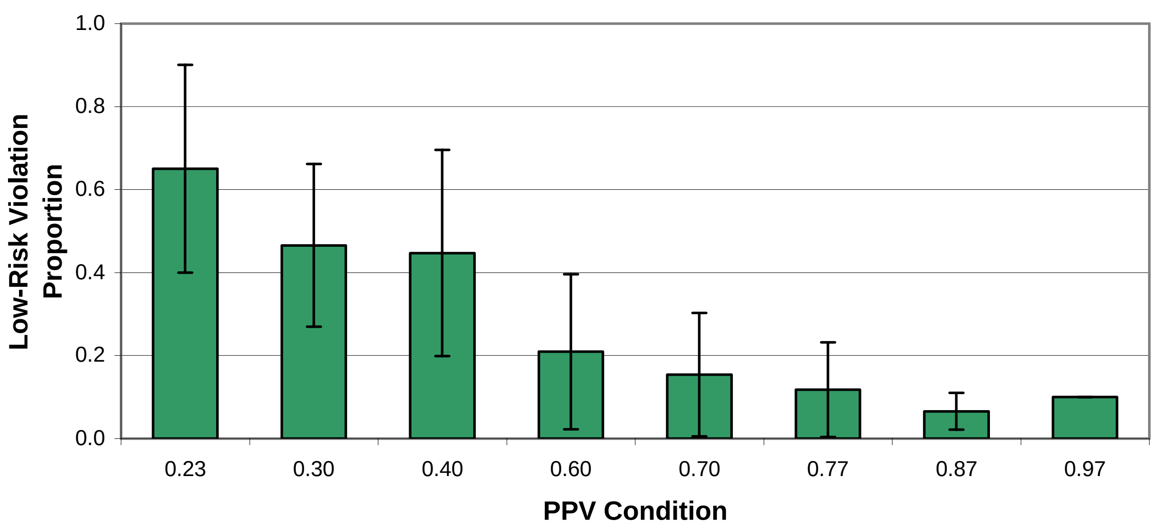


Figure A-2. Means and Standard Deviations for No-Risk Violation Proportions Means

Figure A-3 demonstrates that a proportion of overall gate violations decreased as a function of warning system reliability (PPV). Proportions of overall gate violations were not subjected to significance testing because they are dependent on overall compliance proportion. However, because a significant PPV effect was found on overall compliance proportions (see Figure 10), it is reasonable to assume that an analogous analysis of overall gate violations would produce significant results.

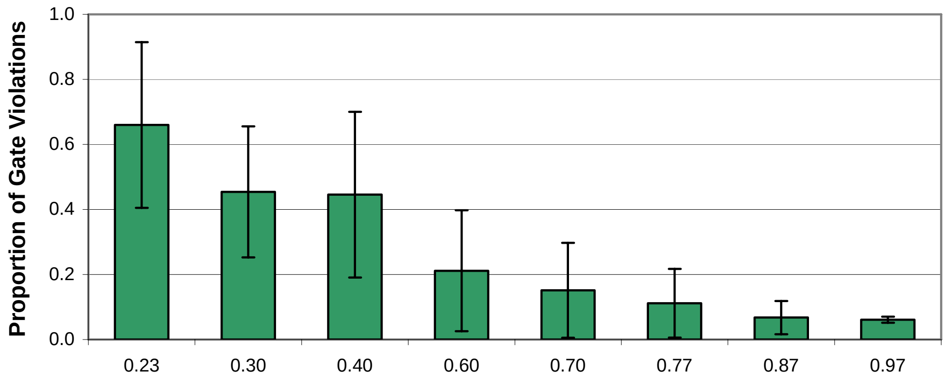


Figure A-3. Proportion of Gate Violations as a Function of Warning System Reliability

APPENDIX B: Grade Crossing Simulator

Highway-Railroad Grade Crossing Simulator Description

<u>Hardware Requirements</u>

The grade crossing simulator is a low fidelity, fixed-based driving simulator that was created with Direct-X Microsoft platform station. The simulator was modeled after a medium-size sedan. The participant uses a Logitech Wingman™ force feedback steering wheel to maneuver the vehicle through a simulated driving course. An accelerator and a break pedal are secured on the floor, in front of the participant. Vehicle controls and the seat are adjusted depending on participant anthropometrics. Participants are seated 15 feet away from the wall-mounted projection screen. The visual image is displayed using a Barco™ projector on an 8 by 10 ft screen positioned directly in front of the simulator. Simulator controls are enclosed in a 6 by 8 ft wooden box to imitate vehicle enclosure. The simulation depicts a rural driving course with 32 actively protected grade crossings. The driving course is 17-mi long.

A Pentium III desktop computer equipped with a 21-inch monitor is used to set up the simulation. Microsoft Windows 2000 Professional operating system is used to launch and execute the simulation program.

<u>Software Requirements</u>

Simulation Parameters. The experimenter can create a variety of situations typically found at actively protected highway-railroad grade crossings. The experimenter can manipulate the following parameters:

1. Vehicle speed: the maximum vehicle speed was set at 55 mph. This was the fastest speed at which the motorist could travel on the simulated course.
2. Train speed: the speed at which the train arrived at each grade crossing was set to 30 mph.
3. Train length: train length is determined by the number of train cars (including engine). A 20 car-long train was used in the simulation. Traveling at 30 mph, the train occupied each grade crossing for 15 seconds.
4. Events at grade crossings: four different types of events were assigned to each grade crossing separately. Grade crossing events included all four possible events as illustrated by the Signal Detection Theory. They include the following:
    a. flashing lights and a lowered gate, train arrives at the crossing
    b. flashing lights and a lowered gate, train does not arrive at the crossing
    c. no flashing lights, a raised gate, train arrives at the crossing
    d. no flashing lights, a raised gate, train does not arrive at the crossing
5. A payoff matrix: was not used in Study 2, however the simulation affords for bonus points associated with costs and benefits to be used.

Measures. Simulation output file includes objective performance measures obtained from each trial. A text file is generated as the participant negotiates the vehicle through the driving course.

Performance data is recorded as the vehicle approaches and drives through each grade crossing. The following measures are recorded:

a. Motorist behavior: stopping behavior is scored at each grade crossing.
b. Collisions: number vehicle collisions with trains and roadside objects.
c. Collision type: collisions where the vehicle hits the train and when the train hits the vehicle are scored as two different types of collisions.
d. Train time to crossing: the time (seconds) it takes for the train to reach the grade crossing as the vehicle is crossing the tracks.
e. Time to contact: the time (seconds) from the time the warning device is activated to the time when the motorist crosses the tracks.
f. Payoff matrix: the point total that the motorist is earning or loosing as he negotiates the driving course.

Participant Interface. Participants navigate through a 12-mi driving course. The driving course includes rural, divided road with 24 actively protected highway-rail grade crossings. An image of a typical grade crossing with activated warning device is shown in Figure B-1.

Figure B-1. An example of an active grade crossing.

A yellow advance warning sign was positioned on the right side of the road, about 100 ft before the highway-rail grade crossing (Figure B-2).

Figure B-2. Highway-rail grade crossing advance warning sign.

Following the advance warning sign, pavement markings were positioned on the approaching lane, in advance of a highway-rail grade crossing. These pavement markings consist of an X, and the letters RR. Pavement markings are demonstrated in Figure B-3.

Figure B-3. Example pavement markings.

A crossbuck sign, as shown in Figure B-4, was installed on the flashing-light signal assembly on the right side of the road, facing approaching traffic.

Figure B-4. Railroad crossing crossbuck sign.

When activated, the warning signal's two red lights mounted in a horizontal line flash alternatively. A schematic of an automatic gate device with the crossbuck sign is shown in Figure B-5.

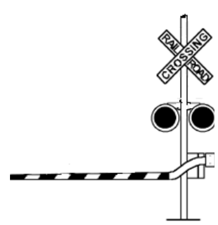


Figure B-5. An automatic gate device with flashing lights.

APPENDIX C: Grade Crossing Events

Table C-1. Grade crossing events experienced by participants in PPV = .40 experimental condition.

Grade Crossing Number	Random Warning Signal Sequence	Grade Crossing Event
1	False Activation	Activated warning signal, train does not arrive
2	False Activation	Activated warning signal, train does not arrive
3	Failure to Activate	Inactivated warning signal, train arrives
4	Proper Inactivation	Inactivated warning signal, train does not arrive
5	False Activation	Activated warning signal, train does not arrive
6	False Activation	Activated warning signal, train does not arrive
7	Failure to Activate	Inactivated warning signal, train arrives
8	Proper Inactivation	Inactivated warning signal, train does not arrive
9	False Activation	Activated warning signal, train does not arrive
10	Proper Inactivation	Inactivated warning signal, train does not arrive
11	Failure to Activate	Inactivated warning signal, train arrives
12	Proper Inactivation	Inactivated warning signal, train does not arrive
13	Failure to Activate	Inactivated warning signal, train arrives
14	Proper Inactivation	Inactivated warning signal, train does not arrive
15	False Activation	Activated warning signal, train does not arrive
16	Proper Activation	Activated warning signal, train arrives
17	Failure to Activate	Inactivated warning signal, train arrives
18	Proper Activation	Activated warning signal, train arrives
19	Failure to Activate	Inactivated warning signal, train arrives
20	Proper Activation	Activated warning signal, train arrives
21	Failure to Activate	Inactivated warning signal, train arrives
22	Proper Activation	Activated warning signal, train arrives
23	False Activation	Activated warning signal, train does not arrive
24	Proper Activation	Activated warning signal, train arrives

Table C-2. Grade crossing events experienced by participants in PPV = .60 experimental condition

Grade Crossing Number	Random Warning Signal Sequence	Grade Crossing Event
1	Proper Activation	Activated warning signal, train arrives
2	Proper Inactivation	Inactivated warning signal, train does not arrive
3	Proper Inactivation	Inactivated warning signal, train does not arrive
4	Proper Activation	Activated warning signal, train arrives
5	Failure to Activate	Inactivated warning signal, train arrives
6	Proper Inactivation	Inactivated warning signal, train does not arrive
7	Proper Activation	Activated warning signal, train arrives
8	Proper Activation	Activated warning signal, train arrives
9	Proper Activation	Activated warning signal, train arrives
10	Proper Inactivation	Inactivated warning signal, train does not arrive
11	False Activation	Activated warning signal, train does not arrive
12	Failure to Activate	Inactivated warning signal, train arrives
13	Proper Inactivation	Inactivated warning signal, train does not arrive
14	Proper Activation	Activated warning signal, train arrives
15	Proper Inactivation	Inactivated warning signal, train does not arrive
16	Failure to Activate	Inactivated warning signal, train arrives
17	Proper Inactivation	Inactivated warning signal, train does not arrive
18	Failure to Activate	Inactivated warning signal, train arrives
19	False Activation	Activated warning signal, train does not arrive
20	False Activation	Activated warning signal, train does not arrive
21	Failure to Activate	Inactivated warning signal, train arrives
22	False Activation	Activated warning signal, train does not arrive
23	Proper Activation	Activated warning signal, train arrives
24	False Activation	Activated warning signal, train does not arrive

Table C-3. Grade crossing events experienced by participants in PPV = .83 experimental condition.

Grade Crossing Number	Random Warning Signal Sequence	Grade Crossing Event
1	Proper Inactivation	Inactivated warning signal, train does not arrive
2	Proper Inactivation	Inactivated warning signal, train does not arrive
3	Proper Activation	Activated warning signal, train arrives
4	Proper Inactivation	Inactivated warning signal, train does not arrive
5	Proper Activation	Activated warning signal, train arrives
6	Proper Activation	Activated warning signal, train arrives
7	Failure to Activate	Inactivated warning signal, train arrives
8	Proper Activation	Activated warning signal, train arrives
9	Proper Inactivation	Inactivated warning signal, train does not arrive
10	False Activation	Activated warning signal, train does not arrive
11	Proper Inactivation	Inactivated warning signal, train does not arrive
12	False Activation	Activated warning signal, train does not arrive
13	Proper Inactivation	Inactivated warning signal, train does not arrive
14	Proper Activation	Activated warning signal, train arrives
15	Proper Activation	Activated warning signal, train arrives
16	Proper Activation	Activated warning signal, train arrives
17	Proper Inactivation	Inactivated warning signal, train does not arrive
18	Proper Inactivation	Inactivated warning signal, train does not arrive
19	Proper Activation	Activated warning signal, train arrives
20	Failure to Activate	Inactivated warning signal, train arrives
21	Proper Inactivation	Inactivated warning signal, train does not arrive
22	Proper Activation	Activated warning signal, train arrives
23	Proper Activation	Activated warning signal, train arrives
24	Proper Inactivation	Inactivated warning signal, train does not arrive

APPENDIX D: Priming Task Results Summary

Priming task data; PPV = .40: probability of hits, false alarms, misses, correct rejections, proportion of correct responses (PCR), sensitivity (d′), and response criterion (λ_c) for each participant.

PPV = .40 Warning System							
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**PCR**	**d'**	λ_c
1	0.99	0.59	0.01	0.41	0.70	2.24	-1.36
2	0.99	0.59	0.01	0.41	0.70	2.24	-1.36
3	0.99	0.60	0.01	0.40	0.70	2.22	-1.36
4	0.75	0.58	0.25	0.42	0.58	0.46	-0.43
5	0.95	0.07	0.05	0.93	0.94	3.18	-0.09
6	0.99	0.59	0.01	0.41	0.70	2.24	-1.36
7	0.99	0.03	0.01	0.97	0.98	4.41	-0.27
8	0.99	0.60	0.01	0.40	0.70	2.22	-1.36
9	0.99	0.58	0.01	0.42	0.71	2.27	-1.34
10	0.65	0.01	0.35	0.99	0.82	2.87	1.04
11	0.99	0.64	0.01	0.36	0.67	1.86	-1.29
12	0.99	0.61	0.01	0.39	0.69	2.20	-1.37
13	0.99	0.59	0.01	0.41	0.70	2.24	-1.36
14	0.99	0.62	0.01	0.38	0.68	1.91	-1.26
15	0.99	0.01	0.01	0.99	0.99	4.95	0.00
16	0.99	0.60	0.01	0.40	0.70	2.22	-1.36
17	0.99	0.59	0.01	0.41	0.70	2.24	-1.36
18	0.99	0.61	0.01	0.39	0.69	2.20	-1.37
19	0.80	0.60	0.20	0.40	0.60	0.59	-0.55
20	0.99	0.57	0.01	0.43	0.71	2.31	-1.32
21	0.45	0.60	0.55	0.40	0.42	-0.39	-0.06
22	0.99	0.57	0.01	0.43	0.71	2.31	-1.32
23	0.99	0.60	0.01	0.40	0.70	2.22	-1.36
24	0.44	0.60	0.56	0.40	0.42	-0.40	-0.05
25	0.96	0.08	0.04	0.92	0.94	3.16	-0.17
Mean	**0.91**	**0.48**	**0.09**	**0.52**	**0.71**	**2.16**	**-0.88**

Priming task data; PPV = .60: probability of hits, false alarms, misses, correct rejections, proportion of correct responses (PCR), sensitivity (d′), and response criterion (λ_c) for each participant.

PPV = .60 Warning System

Subject	pHIT	pFA	pMISS	pCR	PCR	d'	λ_c
1	0.99	0.41	0.01	0.59	0.79	2.71	-1.12
2	0.99	0.39	0.01	0.61	0.80	2.75	-1.10
3	0.99	0.39	0.01	0.61	0.80	2.75	-1.10
4	0.99	0.41	0.01	0.59	0.79	2.45	-0.99
5	0.99	0.10	0.01	0.90	0.94	3.50	-0.47
6	0.99	0.39	0.01	0.61	0.80	2.75	-1.10
7	0.99	0.03	0.01	0.97	0.98	4.41	-0.27
8	0.98	0.39	0.02	0.61	0.79	2.32	-0.89
9	0.99	0.39	0.01	0.61	0.80	2.75	-1.10
10	0.99	0.01	0.01	0.99	0.99	4.69	-0.13
11	0.99	0.39	0.01	0.61	0.80	2.49	-0.97
12	0.99	0.39	0.01	0.61	0.80	2.75	-1.10
13	0.99	0.35	0.01	0.65	0.82	2.85	-1.05
14	0.99	0.39	0.01	0.61	0.80	2.76	-1.09
15	0.99	0.37	0.01	0.63	0.81	2.80	-1.08
16	0.99	0.39	0.01	0.61	0.80	2.75	-1.10
17	0.95	0.39	0.05	0.61	0.78	1.97	-0.70
18	0.99	0.40	0.01	0.60	0.80	2.73	-1.11
19	0.99	0.38	0.01	0.62	0.81	2.78	-1.08
20	0.99	0.01	0.01	0.99	0.99	4.43	0.00
21	0.61	0.41	0.39	0.59	0.60	0.52	-0.03
22	0.99	0.32	0.01	0.68	0.84	2.94	-1.00
23	0.99	0.41	0.01	0.59	0.79	2.71	-1.12
24	0.96	0.01	0.04	0.99	0.97	3.97	0.23
25	0.92	0.04	0.08	0.96	0.94	3.16	0.17
Mean	**0.97**	**0.30**	**0.03**	**0.70**	**0.83**	**2.91**	**-0.77**

Priming task data; PPV = .83: probability of hits, false alarms, misses, correct rejections, proportion of correct responses (PCR), sensitivity (d′), and response criterion (λ_c) for each participant.

PPV = .83 Warning System							
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**PCR**	**d'**	λ_c
s-1	0.99	0.15	0.01	0.85	0.92	3.50	-0.73
s-2	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-3	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-4	0.99	0.13	0.01	0.87	0.93	3.36	-0.54
s-5	0.99	0.02	0.01	0.98	0.98	4.27	-0.08
s-6	0.99	0.15	0.01	0.85	0.92	3.53	-0.71
s-7	0.99	0.01	0.01	0.99	0.99	4.69	-0.13
s-8	0.99	0.17	0.01	0.83	0.91	3.18	-0.62
s-9	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-10	0.99	0.05	0.01	0.95	0.97	4.15	-0.40
s-11	0.99	0.17	0.01	0.83	0.91	3.42	-0.77
s-12	0.99	0.16	0.01	0.84	0.92	3.47	-0.74
s-13	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-14	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-15	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-16	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-17	0.99	0.15	0.01	0.85	0.92	3.50	-0.73
s-18	0.99	0.16	0.01	0.84	0.92	3.47	-0.74
s-19	0.99	0.17	0.01	0.83	0.91	3.44	-0.75
s-20	0.99	0.08	0.01	0.92	0.95	3.62	-0.41
s-21	0.83	0.17	0.17	0.83	0.83	1.91	0.01
s-22	0.99	0.04	0.01	0.96	0.98	4.23	-0.36
s-23	0.99	0.16	0.01	0.84	0.92	3.47	-0.74
s-24	0.85	0.17	0.15	0.83	0.84	1.99	-0.03
s-25	0.96	0.05	0.04	0.95	0.95	3.36	-0.07
Mean	**0.98**	**0.13**	**0.02**	**0.87**	**0.92**	**3.47**	**-0.55**

APPENDIX E: Driving Task Results Summary

Extended matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, and proportion of correct responses obtained in PPV = .40 condition.

PPV = .40 Warning System Proper Function

Subject	pHIT	pFA	pMISS	pCR	PCR
s-1	1.00	0.00	0.00	1.00	1.00
s-2	0.00	0.00	1.00	1.00	0.50
s-3	0.80	0.00	0.20	1.00	0.90
s-4	0.00	0.00	1.00	1.00	0.50
s-5	0.00	0.00	1.00	1.00	0.50
s-6	1.00	0.00	0.00	1.00	1.00
s-7	0.00	0.00	1.00	1.00	0.50
s-8	0.80	0.00	0.20	1.00	0.90
s-9	0.40	0.00	0.60	1.00	0.70
s-10	0.20	0.00	0.80	1.00	0.60
s-11	1.00	0.00	0.00	1.00	1.00
s-12	0.00	0.00	1.00	1.00	0.50
s-13	1.00	0.00	0.00	1.00	1.00
s-14	0.40	0.00	0.60	1.00	0.70
s-15	0.80	0.00	0.20	1.00	0.90
s-16	0.20	0.00	0.80	1.00	0.60
s-17	1.00	0.00	0.00	1.00	1.00
s-18	0.60	0.00	0.40	1.00	0.80
s-19	0.00	0.00	1.00	1.00	0.50
s-20	0.60	0.00	0.40	1.00	0.80
s-21	1.00	0.00	0.00	1.00	1.00
s-22	0.20	0.00	0.80	1.00	0.60
s-23	1.00	0.00	0.00	1.00	1.00
s-24	0.80	0.00	0.20	1.00	0.90
s-25	0.00	0.00	1.00	1.00	0.50
Mean	**0.51**	**0.00**	**0.49**	**1.00**	**0.76**

Extended matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, and proportion of correct responses obtained in PPV = .40 condition.

PPV = .40 Warning System Failure					
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**PCR**
s-1	1.00	1.00	0.00	0.00	0.50
s-2	0.00	0.00	1.00	1.00	0.50
s-3	0.86	0.00	0.14	1.00	0.93
s-4	0.00	0.00	1.00	1.00	0.50
s-5	0.00	0.00	1.00	1.00	0.50
s-6	1.00	0.86	0.00	0.14	0.57
s-7	0.00	0.00	1.00	1.00	0.50
s-8	0.14	0.43	0.86	0.57	0.36
s-9	0.14	0.00	0.86	1.00	0.57
s-10	0.14	0.00	0.86	1.00	0.57
s-11	0.86	1.00	0.14	0.00	0.43
s-12	0.00	0.00	1.00	1.00	0.50
s-13	0.86	0.86	0.14	0.14	0.50
s-14	0.57	0.14	0.43	0.86	0.71
s-15	0.00	0.57	1.00	0.43	0.22
s-16	0.00	0.14	1.00	0.86	0.43
s-17	0.14	0.57	0.86	0.43	0.29
s-18	0.29	0.14	0.71	0.86	0.57
s-19	0.00	0.00	1.00	1.00	0.50
s-20	0.00	0.14	1.00	0.86	0.43
s-21	0.00	0.43	1.00	0.57	0.29
s-22	0.00	0.00	1.00	1.00	0.50
s-23	0.29	0.57	0.71	0.43	0.36
s-24	0.00	0.00	1.00	1.00	0.50
s-25	0.57	0.14	0.43	0.86	0.71
Mean	**0.27**	**0.28**	**0.73**	**0.72**	**0.50**

Extended matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, and proportion of correct responses obtained in PPV = .60 condition.

PPV = .60 Warning System Proper Function					
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**PCR**
s-1	1.00	0.00	0.00	1.00	1.00
s-2	0.00	0.00	1.00	1.00	0.50
s-3	0.71	0.00	0.29	1.00	0.86
s-4	0.71	0.00	0.29	1.00	0.86
s-5	1.00	0.00	0.00	1.00	1.00
s-6	1.00	0.00	0.00	1.00	1.00
s-7	0.00	0.00	1.00	1.00	0.50
s-8	1.00	0.00	0.00	1.00	1.00
s-9	0.14	0.00	0.86	1.00	0.57
s-10	0.29	0.00	0.71	1.00	0.64
s-11	1.00	0.00	0.00	1.00	1.00
s-12	0.00	0.00	1.00	1.00	0.50
s-13	0.43	0.00	0.57	1.00	0.71
s-14	0.29	0.00	0.71	1.00	0.64
s-15	0.86	0.00	0.14	1.00	0.93
s-16	0.29	0.00	0.71	1.00	0.64
s-17	1.00	0.00	0.00	1.00	1.00
s-18	0.86	0.00	0.14	1.00	0.93
s-19	0.29	0.00	0.71	1.00	0.64
s-20	1.00	0.00	0.00	1.00	1.00
s-21	0.71	0.00	0.29	1.00	0.86
s-22	0.43	0.00	0.57	1.00	0.71
s-23	1.00	0.00	0.00	1.00	1.00
s-24	0.86	0.00	0.14	1.00	0.93
s-25	0.14	0.00	0.86	1.00	0.57
Mean	**0.60**	**0.00**	**0.40**	**1.00**	**0.80**

Extended matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, and proportion of correct responses obtained in PPV = .60 condition.

PPV = .60 Warning System Failure

Subject	pHIT	pFA	pMISS	pCR	PCR
s-1	0.80	0.40	0.20	0.60	0.70
s-2	0.00	0.00	1.00	1.00	0.50
s-3	1.00	0.00	0.00	1.00	1.00
s-4	0.20	0.40	0.80	0.60	0.40
s-5	1.00	0.00	0.00	1.00	1.00
s-6	1.00	1.00	0.00	0.00	0.50
s-7	0.00	0.00	1.00	1.00	0.50
s-8	1.00	0.20	0.00	0.80	0.90
s-9	0.00	0.00	1.00	1.00	0.50
s-10	0.20	0.00	0.80	1.00	0.60
s-11	1.00	0.80	0.00	0.20	0.60
s-12	0.00	0.00	1.00	1.00	0.50
s-13	0.60	0.00	0.40	1.00	0.80
s-14	1.00	0.20	0.00	0.80	0.90
s-15	0.00	0.40	1.00	0.60	0.30
s-16	0.00	0.00	1.00	1.00	0.50
s-17	0.40	0.80	0.60	0.20	0.30
s-18	0.20	0.40	0.80	0.60	0.40
s-19	0.00	0.00	1.00	1.00	0.50
s-20	0.20	0.60	0.80	0.40	0.30
s-21	0.00	0.60	1.00	0.40	0.20
s-22	0.00	0.00	1.00	1.00	0.50
s-23	1.00	0.40	0.00	0.60	0.80
s-24	0.80	0.60	0.20	0.40	0.60
s-25	0.00	0.00	1.00	1.00	0.50
Mean	**0.42**	**0.27**	**0.58**	**0.73**	**0.57**

Extended matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, and proportion of correct responses obtained in PPV = .83 condition.

PPV = .83 Warning System Proper Function					
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**PCR**
s-1	1.00	0.00	0.00	1.00	1.00
s-2	0.30	0.00	0.70	1.00	0.65
s-3	1.00	0.00	0.00	1.00	1.00
s-4	0.10	0.00	0.90	1.00	0.55
s-5	0.00	0.00	1.00	1.00	0.50
s-6	1.00	0.00	0.00	1.00	1.00
s-7	0.00	0.00	1.00	1.00	0.50
s-8	1.00	0.00	0.00	1.00	1.00
s-9	0.20	0.00	0.80	1.00	0.60
s-10	0.50	0.00	0.50	1.00	0.75
s-11	0.80	0.00	0.20	1.00	0.90
s-12	0.40	0.00	0.60	1.00	0.70
s-13	1.00	0.00	0.00	1.00	1.00
s-14	0.20	0.00	0.80	1.00	0.60
s-15	1.00	0.00	0.00	1.00	1.00
s-16	0.00	0.00	1.00	1.00	0.50
s-17	1.00	0.00	0.00	1.00	1.00
s-18	0.20	0.00	0.80	1.00	0.60
s-19	0.00	0.00	1.00	1.00	0.50
s-20	0.80	0.00	0.20	1.00	0.90
s-21	1.00	0.00	0.00	1.00	1.00
s-22	1.00	0.00	0.00	1.00	1.00
s-23	1.00	0.00	0.00	1.00	1.00
s-24	0.80	0.00	0.20	1.00	0.90
s-25	0.30	0.00	0.70	1.00	0.65
Mean	**0.58**	**0.00**	**0.42**	**1.00**	**0.79**

Extended matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, and proportion of correct responses obtained in PPV = .83 condition.

PPV = .83 Warning System Failure

Subject	pHIT	pFA	pMISS	pCR	PCR
s-1	1.00	1.00	0.01	0.01	0.50
s-2	0.01	0.01	1.00	1.00	0.50
s-3	1.00	0.50	0.01	0.50	0.75
s-4	0.01	0.01	1.00	1.00	0.50
s-5	0.01	0.01	1.00	1.00	0.50
s-6	1.00	0.01	0.01	1.00	1.00
s-7	0.01	0.01	1.00	1.00	0.50
s-8	1.00	0.50	0.01	0.50	0.75
s-9	0.01	0.01	1.00	1.00	0.50
s-10	0.50	0.01	0.50	1.00	0.75
s-11	0.01	0.01	1.00	1.00	0.50
s-12	0.01	0.01	1.00	1.00	0.50
s-13	1.00	0.01	0.01	1.00	1.00
s-14	0.50	0.01	0.50	1.00	0.75
s-15	1.00	1.00	0.01	0.01	0.50
s-16	0.01	0.01	1.00	1.00	0.50
s-17	1.00	1.00	0.01	0.01	0.50
s-18	0.50	0.01	0.50	1.00	0.75
s-19	0.01	0.01	1.00	1.00	0.50
s-20	0.01	0.01	1.00	1.00	0.50
s-21	0.01	0.01	1.00	1.00	0.50
s-22	1.00	0.01	0.01	1.00	1.00
s-23	0.01	0.01	1.00	1.00	0.50
s-24	0.01	0.50	1.00	0.50	0.25
s-25	0.50	0.01	0.50	1.00	0.75
Mean	**0.40**	**0.18**	**0.60**	**0.82**	**0.61**

Singe matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, sensitivity and response bias measures obtained in PPV = .40 condition.

PPV = .40 Warning System						
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**d'**	**λ-center**
s-1	1.00	0.58	0.00	0.42	2.93	-1.68
s-2	0.00	0.00	1.00	1.00	0.00	3.14
s-3	0.83	0.00	0.17	1.00	4.11	1.09
s-4	0.00	0.00	1.00	1.00	0.00	3.14
s-5	0.00	0.00	1.00	1.00	0.00	3.14
s-6	1.00	0.50	0.00	0.50	3.14	-1.57
s-7	0.00	0.00	1.00	1.00	0.00	3.14
s-8	0.42	0.25	0.58	0.75	0.46	0.44
s-9	0.25	0.00	0.75	1.00	2.47	1.91
s-10	0.17	0.00	0.83	1.00	2.18	2.06
s-11	0.92	0.58	0.08	0.42	1.17	-0.80
s-12	0.00	0.00	1.00	1.00	0.00	3.14
s-13	0.92	0.50	0.08	0.50	1.38	-0.69
s-14	0.50	0.08	0.50	0.92	1.38	0.69
s-15	0.33	0.33	0.67	0.67	0.00	0.43
s-16	0.08	0.08	0.92	0.92	0.00	1.38
s-17	0.50	0.33	0.50	0.67	0.43	0.22
s-18	0.42	0.08	0.58	0.92	1.17	0.80
s-19	0.00	0.00	1.00	1.00	0.00	3.14
s-20	0.25	0.08	0.75	0.92	0.71	1.03
s-21	0.42	0.25	0.58	0.75	0.46	0.44
s-22	0.08	0.00	0.92	1.00	1.76	2.26
s-23	0.58	0.33	0.42	0.67	0.64	0.11
s-24	0.33	0.00	0.67	1.00	2.71	1.79
s-25	0.33	0.08	0.67	0.92	0.95	0.91
Mean	**0.37**	**0.16**	**0.63**	**0.84**	**1.12**	**1.19**

Singe matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, sensitivity and response bias measures obtained in PPV = .60 condition.

PPV = .60 Warning System						
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**d'**	**λ-center**
s-1	0.92	0.17	0.08	0.83	2.35	-0.21
s-2	0.00	0.00	1.00	1.00	0.00	3.14
s-3	0.83	0.00	0.17	1.00	4.11	1.09
s-4	0.50	0.17	0.50	0.83	0.97	0.48
s-5	1.00	0.00	0.00	1.00	6.29	0.00
s-6	1.00	0.42	0.00	0.58	3.36	-1.47
s-7	0.00	0.00	1.00	1.00	0.00	3.14
s-8	1.00	0.08	0.00	0.92	4.53	-0.88
s-9	0.08	0.00	0.92	1.00	1.76	2.26
s-10	0.25	0.00	0.75	1.00	2.47	1.91
s-11	1.00	0.33	0.00	0.67	3.57	-1.36
s-12	0.00	0.00	1.00	1.00	0.00	3.14
s-13	0.50	0.00	0.50	1.00	3.14	1.57
s-14	0.58	0.08	0.42	0.92	1.59	0.59
s-15	0.50	0.17	0.50	0.83	0.97	0.48
s-16	0.17	0.00	0.83	1.00	2.18	2.06
s-17	0.75	0.33	0.25	0.67	1.11	-0.12
s-18	0.58	0.17	0.42	0.83	1.18	0.38
s-19	0.17	0.00	0.83	1.00	2.18	2.06
s-20	0.67	0.25	0.33	0.75	1.11	0.12
s-21	0.42	0.25	0.58	0.75	0.46	0.44
s-22	0.25	0.00	0.75	1.00	2.47	1.91
s-23	1.00	0.17	0.00	0.83	4.11	-1.09
s-24	0.83	0.25	0.17	0.75	1.64	-0.15
s-25	0.08	0.00	0.92	1.00	1.76	2.26
Mean	**0.52**	**0.11**	**0.48**	**0.89**	**2.13**	**0.87**

Singe matrix driving task data for each participant. Probability of hits, false alarms, misses, correct rejections, sensitivity and response bias measures obtained in PPV = .83 condition.

PPV = .83 Warning System						
Subject	**pHIT**	**pFA**	**pMISS**	**pCR**	**d'**	**λ-center**
s-1	1.00	0.17	0.00	0.83	4.11	-1.09
s-2	0.25	0.00	0.75	1.00	2.47	1.91
s-3	1.00	0.08	0.00	0.92	4.53	-0.88
s-4	0.08	0.00	0.92	1.00	1.76	2.26
s-5	0.00	0.00	1.00	1.00	0.00	3.14
s-6	1.00	0.00	0.00	1.00	6.29	0.00
s-7	0.00	0.00	1.00	1.00	0.00	3.14
s-8	1.00	0.08	0.00	0.92	4.53	-0.88
s-9	0.17	0.00	0.83	1.00	2.18	2.06
s-10	0.50	0.00	0.50	1.00	3.14	1.57
s-11	0.67	0.00	0.33	1.00	3.57	1.36
s-12	0.33	0.00	0.67	1.00	2.71	1.79
s-13	1.00	0.00	0.00	1.00	6.29	0.00
s-14	0.25	0.00	0.75	1.00	2.47	1.91
s-15	1.00	0.17	0.00	0.83	4.11	-1.09
s-16	0.00	0.00	1.00	1.00	0.00	3.14
s-17	1.00	0.17	0.00	0.83	4.11	-1.09
s-18	0.25	0.00	0.75	1.00	2.47	1.91
s-19	0.00	0.00	1.00	1.00	0.00	3.14
s-20	0.67	0.00	0.33	1.00	3.57	1.36
s-21	0.83	0.00	0.17	1.00	4.11	1.09
s-22	1.00	0.00	0.00	1.00	6.29	0.00
s-23	0.83	0.00	0.17	1.00	4.11	1.09
s-24	0.67	0.08	0.33	0.92	1.81	0.48
s-25	0.33	0.00	0.67	1.00	2.71	1.79
Mean	**0.55**	**0.03**	**0.45**	**0.97**	**3.09**	**1.12**

References

Abraham, J., Datta, T.K., and Datta, S. (1998). Motorist behavior at rail-highway crossings. *Transportation Research Record, 1648*, 28–34.

Benekohal, R.F. and Aycin, M. F. (2004). *Analyses of Motorists' Opinions about Railroad Grade Crossings Traffic Control Devices and Safety: Background Survey* (FHWA-IL/UI-TOL-10). Urbana, IL: University of Illinois at Urbana-Champaign.

Benekohal, R.F. and Aycin, M. F. (2002). *Performance Evaluation of the Pilot Study of Advisory On-board Vehicle Warning Systems at Railroad Grade Crossings* (FHWA-IL/UI-TOL-4). Urbana, IL: University of Illinois at Urbana-Champaign.

Benekohal, R.F. and Rawls, C.G. (2004a). *Analyses of the Motorists' Responses in Final Surveys to the In-Vehicle Receiver (IVR)* (FHWA-IL/UI-TOL-13). Urbana, IL: University of Illinois at Urbana-Champaign.

Benekohal, R.F. and Rawls, C. G. (2004b). *Analyses of the Motorists' Responses (in Survey Number 2) to the In-Vehicle Receiver (IVR) After Experiencing One Mode of Operation* (FHWA-IL/UI-TOL-11). Urbana, IL: University of Illinois at Urbana-Champaign.

Benekohal, R.F. (2004). Findings of Focus Group Meetings for the Pilot Study of Advisory On-Board Vehicle Warning Systems at Railroad Grade Crossings (FHWA-IL/UI-TOL-14). Urbana, IL: University of Illinois at Urbana-Champaign.

Bitan, Y., Meyer, J., Shinar, D., and Zmora, E. (2000). Staff actions and alarms in a neonatal intensive unit. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (Vol. 1, pp.17–20). Santa Monica, CA.

Bliss, J.P. and Gilson, R.D. (1998). Emergency signal failure: implications and recommendations. *Ergonomics, 41*(1), 57–72.

Bliss, J.P., Gilson, R.D., and Deaton, J.E. (1995). Human probability matching behavior in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300–2312.

Breznitz, S. (1983). *Cry wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brunswik E. (1937). Psychology as a science of objective relations. *Philosophy of Science, 4*, 227–260.

Brunswik E. (1952). *The conceptual framework of psychology.* Chicago: University of Chicago Press, *Int. Encycl. unified Sci.,* Vol. 1, No. 10.

Caird, J.K., Creaser, J.I., Edwards, C.J., and Dewar, R.E. (2002). A human factors analysis of highway-railway grade crossing accidents in Canada. Report No. TP 13938E. Transport Canada Publication. Retrieved from http://www.tc.gc.ca/tdc/publication/pdf/13900/13938e.pdf

Carroll, A.A. and Haines, M. (2002). *The use of photo enforcement at highway-rail grade crossings in the U.S*. Transportation Safety Board [CD-ROM]. Washington, DC: TRB.

MotoristCooksey, R.W. (1996). *Judgment analysis: theory, methods, and applications*, New York, Academic Press, Inc.

Chugh, J.S. and Caird, J.K. (1999). In-vehicle train warnings (ITW): The effect of reliability and failure type on motorist perception response time and trust. *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society Meeting* (pp. 1012–1016). Santa Monica, CA: Human Factors and Ergonomics Society.

Cotté, N., Meyer, J., and Coughlin, J. F. (2001). Older and younger motorist's reliance on collision warning systems. Proceedings of the 45th Annual Meeting of the Human Factor Society (pp. 277–-280). Santa Monica, CA: Human Factors and Ergonomics Society.

Dixon, S.R., Wickens, C.D., and Chang, D. (2004). Unmanned Aerial Vehicle flight control: false alarms versus misses. *Proceedings of the Human Factors and Ergonomics Society*. Santa Monica, CA.

Dow, S., Thomas, G., and Johnson, L. (1999). *Signal detection performance with a haptic device*. *Proceedings of the Human Factors and Ergonomics Society*. Houston, TX.

Dunn, M. (1995). *Primary-task and secondary-task workload and the cry-wolf phenomenon*. Unpublished doctoral dissertation, University of Alabama, at Huntsville, Huntsville, AL.

Edworthy, J. (2000). *An integrative approach to warnings research*. Paper presented at the 44th annual meeting of Human Factors and Ergonomics Society, San Diego, CA.

Edworthy, J., and Adams, A.S. (1996). *Warning design: a research perspective*. London: Taylor and Francis.

Edworthy, J., and Dale, S. (2000). *Extending knowledge of the effects of social influence in warning compliance*. Paper presented at the 44th annual meeting of Human Factors and Ergonomics Society, San Diego, CA.

Edworthy, J., Loxley, S., and Dennis, I. (1991). Improving auditory warning design: relationship between warning sound parameters and perceived urgency, *Human Factors, 33*, 205–232.

Egan, J.P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Epstein, S. (1990). Cognitive-experiential self-theory. In L. Pervin (Ed.), *Handbook of personality theory and research* (pp. 165–192). New York: Guilford Press.

Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71*(2), 390–-405.

Federal Highway Administration (2002). *Guidance on Traffic Control Devices at Highway-Rail Grade Crossings*. Washington, D.C.: U.S. Department of Transportation.

Getty, D.J., Swets, J.A., Pickett, R.M., and Gonthier, D. (1995). System operator response to warnings of danger: a laboratory investigation of the effects of the predictive value of a warning on a response time. *Journal of Experimental Psychology: Applied, 1*(1), 19–33.

Gilson, R.D., Mouloua, M., Graft, A.S., and McDonald, D.P. (2001). Behavioral influences of proximal alarms. *Human Factors, 43*(4), 595–610.

Green, D.M., and Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gossard, W. (1987). Sight distance: A factor in highway-rail grade crossing accidents (pp.16–-19). *Proceedings of the, 1987 National Conference on Highway-Rail Safety,* Denver, CO, September.

Hauer, E., and Persaud, B.N. (1986). Rail-highway grade crossings: their safety and the effect of warning devices. *Proceedings of the 30th Annual American Association for Automotive Medicine* (pp. 247–262). Montreal, QC.

Heathington, K. W. (1996). Railroad grade crossing accident behavior and countermeasures. *Infrastructure, 2*, 10–16.

Heathington, K.W., and Urbanik, T. (1972). *Motorist information systems for highway-railway grade crossings* (pp. 59–77). Highway Research Record, 414. Washington, DC Highway Research Board.

Holzworth, R. J. (2001). Judgment analysis. In K. R. Hammond and T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 324–-327). New York: Oxford University Press.

Janis, I.L. (1962). Psychological effects of warnings. In G. W. Baker and D. W. Chapman (Eds.), *Man and Society in Disaster*. New York: Basic Books.

Jentsch, F., Bowers, C., Compton, D., Navarro, G., and Tait, T. (1996). Team-track: A tool for investigating tracking performance in teams. *Behavior Research Methods*, *Instruments, & Computers, 28*(3), 411–417.

Jurgenson, T., Park, I., Sheridan, T.B., and Meyer, J. (2001). *Conflict warnings and the search and detection of collisions in a simulated ATC-like task*. Paper presented as the 44th meeting of Human Factors and Ergonomics Society, San Diego, CA.

Keppel, G. (1973). *Design and analysis: a researcher's handbook* (3rd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Kerr, J.L. (1985). Auditory warnings in intensive care units and operating theatres. *Ergonomics International, 85*, 172–174.

Kirkpatrick, L. and Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology, 63*(4), 534–544.

Lee, J.D., and Moray, N. (1994). Trust, self-confidence, and operator's adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153–184.

Lee, J.D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*, (1), pp. 50–80.

Leibowitz, H.W. (1985). Grade crossing accidents and human factors engineering. *American Scientist, 73*, 558–563.

Lehto, M.R. (2006). Optimal Warnings: An Information and Decision Theoretic Perspective. In M.S. Wogalter (Ed.), *Handbook of Warnings*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lerner, N., Ratte, D., and Walker, J. (1990). *Motorist behavior at rail-highway crossings*. Report No. FHWA-SA-90-008. Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Mallett, L.G., Vaught, C., and Brnich, M.J., Jr. (1993). Sociotechnical communication in an underground mine fire: a study of warning messages during an emergency evacuation. *Safety Science, 16*, 709–728.

Macmillan, N. A., and Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Maltz, M., and Meyer, J. (2001). Use of warnings in an attentionally demanding detection task. *Human Factors, 43*(2), 217–226.

Maltz, M. and Shinar, D. (2000). *Human factors in cued target acquisition*. Paper presented at the 44th meeting of Human Factors and Ergonomics Society, San Diego, CA.

Maltz, M., and Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *Human Factors, 45*, 281–295.

Maltz, M., and Shinar, D. (2004). Imperfect in-vehicle collision avoidance warning systems can aid motorists. *Human Factors*, *46*(2), 357–366.

Meeker, F., Fox, D., and Weber, C. (1997). A comparison of motorist behavior at railroad grade crossings with two different protection systems. *Accident Analysis and Prevention*, 29(1), 11–16.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors, 43*(4), 563–572.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.

Meyer, J. and Bitan, Y. (2002). Why better operators receive worse warnings. *Human Factors, 44*(3), 343–353.

Morrissey, J. (1980). *The effectiveness of flashing lights and flashing lights with gates in reducing accident frequency at public rail-highway crossings* (Report No. FRA-RRS-80-005). Waltham MA: Input Output Services.

Mosier, K.L., Skitka, L.J., Heers, S., and Burdick, M.D. (1998). Automation bias: decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 8*, 47–63.

Muir, B.M. and Moray, N. (1996). Trust in automation: Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*, 429–460.

Multer, J., Conti, J., and Sheridan, T. (2000). *Recognition of rail car retroreflective patterns for improving nighttime conspicuity*. Report No. DOT/FRA/ORD-00/07. Washington, D.C.: U.S. Department of Transportation, Federal Railroad Administration.

Nakata, A. and Noel, R.W. (2001). *Intelligibility of vocoded and waveform speech for native and non-native speakers of English*. Paper presented at the 45th meeting of Human Factors and Ergonomics Society, Minneapolis, MN.

National Transportation Safety Board (1998). *Safety study: Safety at passive grade crossings, Volume 2: Analysis* (Report No. PB98-917004, NTSB/SS-98/02). Washington, DC: NTSB.

Pacini, R., and Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*(6), 972–987.

Parasuraman, R., Masalonis, A.J., and Hancock, P.A. (2000). Fuzzy signal detection theory: basic postulates and formulas for analyzing and machine performance. *Human Factors, 42*(4), 636–659.

Parasuraman, R., Molloy, R., and Singh, I.L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology, 3*, 1–23.

Parasuraman, R. and Riley, V. (1997). Humans and automation: use, misuse, disuse, and abuse. *Human Factors, 39*, 230–253.

Parker, D., Reason, J. T., Manstead, A. S. R., and Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics, 38*(5), 1036–1048.

Pate-Cornell, M.E. (1996). Warning systems in risk management. *Risk Analysis, 6*(2), 223–234.
Patterson, R.D. (1982). *Guidelines for auditory systems on civil aircraft*. Civil aviation authority paper 82017, London, UK.

Peterson, D.A., Uhlarik, J., Raddatz, K.R., and Ward, J. (1999). *The role of situational factors and display cues on collision detection and avoidance using a terrain enhanced primary flight display*. Paper presented at the 43rd meeting of Human Factors and Ergonomics Society, Houston, TX.

Raab, F.H., Brooker, M.C., Ryan, T.E., and Weachter, J.R. (1977). *Innovative concepts and technology for railroad-highway grade crossing motorist warning systems*. Vol. I. Overview and concept generation and analysis (Report No. FRA/ORD-77/37.I) Cambridge, Massachusetts: US Department of Transportation, Transportation Systems Center.

Raslear, T.G. (1996). Motorist behavior at rail-highway grade crossings: A signal detection theory analysis. In A. A. Carroll and J. L. Helser (Eds.), *Safety of highway-railroad grade crossings*. Research needs workshop. Volume II – Appendices (Report No. DOT/FRA/ORD-95/14.2, DOT-VNTSC-FRA-95-12.2, pp. F9-F56). Washington, DC: U.S. Department of Transportation.

SRF Consulting Group, Inc. (1998). *In-vehicle signing for school buses at railroad-highway grade crossings. Evaluation Report*. (SRF No. 0972870). Minneapolis, MN: Minnesota Department of Transportation.

Sanders, M.S. and McCormick, E.J. (1993). *Human factors in engineering and design*. New York: McGraw-Hill.

Shinar, D. and Raz, S. (1982). Motorist response to different railroad crossing protection systems. *Ergonomics, 25*(9), 801–808.

Shulte, W.R. (1975). Effectiveness of automatic warning devices in reducing accidents at grade crossings. *Transportation Research Record, 611*, 49–57.

Sorkin, R.D. (1988). Why are people turning off our alarms? *Acoustical Society of America, 84* (3), 1107–1108.

Sorkin, R.D. and Woods, D.D. (1985). Systems with human monitors: a signal detection analysis. *Human-Computer Interaction, 1*, 49-75.

U.S. Federal Highway Administration. (2000). *Railroad safety statistics*. Washington, DC: U.S. Department of Transportation, FRA. (http://safetydata.fra.dot.gov/)

U.S. Federal Highway Administration. (2003). *Manual on Uniform Traffic Control Devices (MUTCD*). http://mutcd.fhwa.dot.gov/ (2005, April 1)

Walker, F.W. and Roberts, S.E. (1975). *Influence of lighting on accident frequency at highway intersection*. Ames, IA: Department of Transportation.

Wickens, T.D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.

Wigglesworth, E. C. (1987). A human factors commentary on innovations at railroad-highway grade crossings in Australia. *Journal of Safety Research, 32*, 309–321.

Wilde, G.J.S., Hay, M.C., and Brites, J.N. (1987). *Video-recorded motorist behaviour at railway crossings: Approach speeds and critical incidents* (Canadian Institute for Guided Ground Transport Report No. 87-6/Transport Canada Report No. TP-9014E). Montreal, Quebec: Transportation Development Centre Transport Canada.

Yeh, M. and Multer, J. (in preparation). *Motorist Behavior at Highway-Railroad Grade Crossings: A Literature Review from 1990–2006.*