Bridging laboratory and field studies

Jan-Erik Källhammer & Kip Smith

The behavioral research community has long acknowledged the need for methods that capture both the rigor of the laboratory and the ecological validity of the field (Brehmer & Dörner, 1993). The method we present here - retrospective review and rating of field data - is designed to bridge the lab and the field. It is tailored for studies of driver acceptance of active safety systems but is not limited to that application.

The method presents to observers a set of video recordings of events captured during a field operational test (FOT). We use a prototype active safety system with a reasonable level of performance to flag events in the field. Thus, the flagged events are all candidates for alerts from the active safety system. The events can be crashes, 'incidents', 'near-incidents' or, indeed, any situation of interest. The method treats the recorded events as experimental stimuli. The good quality of the FOT recordings retain much of the ecological validity of actual traffic events. Fully situated contextualization is, of course, achieved only in the moment.

Back in the lab, we review the flagged events and categorize them along dimensions of theoretical importance or pragmatic interest. In a second review, we select a subset of events that that addresses the needs of the system's developers and that is likely to span the range of driver acceptance. The criterion for selection of clips to be reviewed was that there should not be any ambiguity regarding which pedestrian(s) the volunteers were to rate. Groups of pedestrians were allowed, if they were in the same context, for example walking together. Clips with pedestrians visible at different locations were excluded. Bicyclists were also excluded, as the perception of risk related to them may have been different and their patterns of motion are different than that of pedestrians.

The product of this selection process is a series of video clips of traffic incidents and situations that is amenable to replay and systematic analysis. Each clip is typically 20+ seconds long to provide observers with a sense of the traffic context. The categorization process and randomizing the order of presentation contribute experimental rigor to the review.

The experimental procedure consists of viewing and rating: Observers in the laboratory watch the replay of an event and then rate the level with which they would likely accept an alert from an active safety system to that event. The rating procedure was inspired by van der Laan, Heino & De Waard (1997). The rating scale typically ranges from 'totally unacceptable' to 'totally acceptable'. In some studies we have used a slider bar for a continuous measure (Källhammer, Smith, Karlsson, & Hollnagel, 2007). In others we have used a Likert-type scale with a sufficient number of levels to allow separation of similar events (Smith & Källhammer, 2010). We typically elicit responses from a large number of observers, leveraging the cost of the FOT and providing sample sizes that are amenable to statistical tests of significance.

To illustrate the viability of the method, we present results from a study of events flagged by a pedestrian warning system. A cross-plot and regression analysis of the ratings of the observers in the laboratory and those of the drivers who experienced the events in traffic reveal a satisfactory level of convergence between the lab and field data. In this study, the retrospective method was used to elicit ratings from 40 observers. The agreement between the ratings by the observers and the drivers is linear and quite good, $r^2 = 0.62$, $F(1,57) = 93.$, $p < .001$. The slope of the regression line is less than 1.0, reflecting the expected regression to the mean. The ratings from the retrospective observers are less extreme than those of the drivers who experienced the events. This finding supports the contention that the method of retrospective review and rating produces data that align with the experience of drivers in the field.

We offer our retrospective review and rating method as a cost-effective approach to bridging the laboratory and the field. It produces reliable and reproducible data that is informing our system design and development.

References

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Computers in Human Behavior, 9 (2-3), 171.

Källhammer, J.-E., Smith, K., Karlsson, J., & Hollnagel, E. (2007). Shouldn't the car react as the driver expects? Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Stevenson, WA.

Smith, K. & Källhammer, J.-E. (June, 2010). Driver acceptance of false alarms to simulated encroachment. Human Factors.

van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research - Part C: Emerging Technologies, 5, 1-10.

Jan-Erik Källhammer
Autoliv Development AB, Vårgårda, Sweden

Kip Smith
CE & DM, Inc. Des Moines, Washington

Jan-Erik Källhammer
Autoliv Research
Autoliv Development AB
SE-447 83 Vårgårda, Sweden
tel: +46 708 320934 (cell)
fax: +46 322 620118
email: jan-erik.kallhammer@autoliv.com

Kip Smith
Cognitive Engineering and Decision Making, Inc.
319 South 210th Street
Des Moines WA 98198 USA
tel: +1 206 824 1512
fax: +46 322 620118
email: kip51smith@gmail.com