Accident Analysis and Prevention 76 (2015) 25–33

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety

William J. Horrey[a,*], Mary F. Lesch[b], Eve Mitsopoulos-Rubens[c], John D. Lee[d]

[a] Liberty Mutual Research Institute for Safety, 71 Frankland Road, Hopkinton, MA 01748, USA
[b] Liberty Mutual Research Institute for Safety, 71 Frankland Road, Hopkinton, MA 01748, USA
[c] VicRoads, 60 Denmark Street, Kew, 3101 VIC, Australia
[d] Department of Industrial and Systems Engineering, University of Wisconsin, 1513 University Avenue, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form 16 December 2014
Accepted 16 December 2014
Available online 2 January 2015

Keywords:
Calibration
Driving
Information selection and utilization
Lens model
Distraction
Automation
Young/novice drivers

ABSTRACT

Humans often make inflated or erroneous estimates of their own ability or performance. Such errors in calibration can be due to incomplete processing, neglect of available information or due to improper weighing or integration of the information and can impact our decision-making, risk tolerance, and behaviors. In the driving context, these outcomes can have important implications for safety. The current paper discusses the notion of calibration in the context of self-appraisals and self-competence as well as in models of self-regulation in driving. We further develop a conceptual framework for calibration in the driving context borrowing from earlier models of momentary demand regulation, information processing, and lens models for information selection and utilization. Finally, using the model we describe the implications for calibration (or, more specifically, errors in calibration) for our understanding of driver distraction, in-vehicle automation and autonomous vehicles, and the training of novice and inexperienced drivers.

1. Introduction

As hungry processors of information, we humans selectively attend to environmental cues and render judgments concerning the state of the world around us. At the same time and with some degree of introspection, we carry out self-appraisals, evaluating how skillful or capable we are in different contexts. Both the manner in which we perceive the world, as well as our perceptions of our own effectiveness as agents in the world, can have a tremendous bearing on the decisions we make, the behaviors we engage in, and the risks we entertain. While these perceptions of world and self often lead to reasonable decisions and behaviors as well as tolerable levels of risk, the more profound interest emerges in situations where the subjective perception of the world or of self deviates from objective reality, with potentially negative consequences. For example, a driver, thinking that current driving conditions are easy, sends a text message—having failed to notice a nearby hazard; the driver with a highly reliable automated system comes to distrust and disuse it for what the driver wrongly considers to be too many false alarms; the novice driver, overconfident in their driving skills and abilities, travels at a high speed on a slippery surface.

Gaps between perception and reality can be related to the notion of calibration, which itself can be broadly defined as the determination of the accuracy of an instrument by measurement of its variation from a standard. Despite having a strong foundation in the physical sciences, calibration is not unique to this context. The concept of calibration also has prominence in the social and psychological sciences—often in situations where one's ability to make sufficiently accurate judgments to guide decision-making and behavior is paramount. Indeed, calibration has been widely discussed in a variety of domains, including weather forecasting, education, scholastic aptitude, medicine, work and management, eye witness testimony, as well as driving. Within these fields, some researchers have focused on whether or not individuals' perceptions are in (or out of) alignment with an objective standard while others have focused on managing task demands and user capabilities as a means of bringing two measurements or estimates into alignment (e.g., Kuiken and Twisk, 2001; Fuller, 2005; Mitsopoulos-Rubens, 2010).

In the current paper, we discuss calibration as it relates to general self-appraisals and evaluations of one's competence in making perceptual judgments (the state of calibration) as well as to models of

* Corresponding author. Tel.: +1 508 497 0237; fax: +1 508 435 0482.
E-mail addresses: william.horrey@libertymutual.com (W.J. Horrey), mary.lesch@libertymutual.com (M.F. Lesch), eve.mitsopoulosrubens@roads.vic.gov.au (E. Mitsopoulos-Rubens), jdlee@engr.wisc.edu (J.D. Lee).

http://dx.doi.org/10.1016/j.aap.2014.12.017
0001-4575/

self-regulation in driving (which are concerned with the process of bringing subjective and objective measures into alignment). We further propose a framework for calibration in the driving context, building and elaborating upon some earlier models of momentary demand regulation (e.g., Fuller, 2005; Mitsopoulos-Rubens, 2010), models of information processing (e.g., Wickens and Hollands, 2000), and lens models for information selection and utilization (e.g., Brunswik, 1955; Hammond, 1955). We illustrate how different components in the model can account for some of the errors and deficiencies in calibration observed in the literature and, in so doing, overcome possible shortcomings in some previous models of calibration in driving. Finally, we describe the practical implications of calibration for road safety through three examples: our understanding of driver distraction, in-vehicle automation and autonomous vehicles, and the training of inexperienced drivers. It is hoped that the framework will provide guidance to research efforts concerning the role of calibration in the study of road safety; help account for discrepancies between perception, performance and skill; and lead to approaches aimed at mitigating potentially adverse effects of miscalibration.

2. Calibration failures

Errors in calibration have important implications for safety and performance and can be due to deficiencies in the processing of available information, errors in evaluated self-competence and/or comparison errors. The failure to process, or appropriately weigh, highly critical information can result in a slanted, narrow, or erroneous awareness of the situation (e.g., Griffin and Tversky, 1992). Similarly, an unrealistic appraisal of our own skills and abilities—while promoting good feelings of self-worth and esteem—can also place us in situations that we are ill-equipped to deal with.

Evidence from various domains suggests that individuals' subjective impressions or evaluations are not well-calibrated to more objective measures. People generally tend to view themselves in favorable or optimistic terms—regardless of the degree of truth in the assertion (e.g., Brown, 1986; DeJoy, 1989; Dunning et al., 2004). This tendency has been couched in many different terms, such as optimism bias, self-enhancement bias, illusory superiority, among many others (e.g., Sharot, 2011; Hoorens, 1993). Studies of self-efficacy, self-appraisal, self-confidence, and other forms of self-evaluation are widespread in the social and psychological scientific literature. Indeed, there is such a plethora of work in this area that it can even support meta-syntheses of dozens of meta-analyses (e.g., Zell and Krizan, 2014). Findings of enhanced self-appraisals are evident in almost every discipline, including education and learning (e.g., Bol and Hacker, 2012), ethics (e.g., Baumhart, 1968), health and medicine (e.g., Larwood, 1978; Weinstein, 1980; Dunning et al., 2004) and workplace and managerial skills (e.g., Larwood and Whittaker, 1977). This is also the case for self-appraisals with respect to driving skills and abilities. Many studies have documented drivers' tendency to rate themselves more favorably than other drivers or to rate their skills as better than indicated by some objective standard (e.g., Svenson, 1981; McKenna et al., 1991; DeJoy, 1989; Brown and Groeger, 1988; Horswill et al., 2004).[1]

Errors in calibration have also been widely documented in decision-making and judgment paradigms, particularly where uncertainty is a critical element. Calibration, in this framework, is often characterized mathematically as the correspondence between one's confidence in a judgment (degree of certainty) and the accuracy of the judgment (e.g., Murphy, 1973; Lichtenstein and Fischhoff, 1977; Lichtenstein et al., 1982; Baranski and Petrusic, 1994, 1995; Soll, 1996).

Dunning et al. (2004) reviewed evidence for imprecise or flawed self-assessments, citing poor correlations between self-ratings of skill and actual performance as well as the tendency for people to overrate themselves (placing themselves as better than average)—outcomes echoed by many (e.g., Zell and Krizan, 2014; Mabe and West, 1982). These latter ratings are sometimes manifest as overestimates in their engagement in desirable behaviors or in their achievement of favorable outcomes, overly optimistic estimates of future productivity, and over-confidence in judgments. Dunning et al. (2004) further suggested that these errors in calibration can result from incomplete information, on the one hand, and neglect of relevant information, on the other (i.e., in the diligent processing and appropriate weighing of information). Errors can likewise result from incomplete or inadequate processing of available information (through heuristics or selection), biases or other top-down influences (e.g., Slovic et al., 1977; Kahneman and Tversky, 1972).

As hinted in Section 1, our own self-appraisals and evaluations can affect how we interact with the world. Yet many studies do not examine the roles of action or feedback in such evaluations (cf., Simons, 2013; Stone and Opel, 2000). The next section describes how the process of bringing different measures or estimates into alignment has been implemented in models of driving behavior. While "state" errors in calibration of self-evaluation and ability, such as those described above, are inherent to these models, the focus is on the process of appraisal and action regulation (i.e., self-regulation or momentary demand regulation; e.g., Kuiken and Twisk, 2001). Given the dynamic and self-paced nature of driving as well as the lack of a clear gold standard for driving safety/performance, one could argue that the regulatory process is more germane in the driving context than general appraisals of self and skill (although both are important).

2.1. Theories of demand regulation in driving

Several models of driver behavior treat calibration as a regulatory process in which the driver balances the momentary assessment of ability and the assessment of demand. For example, the task-capability interface model (Fuller, 2005) posits that drivers will adopt a preferred level of driving difficulty. Whenever task difficulty exceeds the preferred range, drivers will make behavioral adjustments to return difficulty to the desired range. Difficulty is a property that emerges from the interaction between driving demands and driver capacity (see similar models described by Davidse et al., 2010; de Craen, 2010; Mitsopoulos-Rubens, 2010). Driving demands are determined in part by speed, road, environment, and other driving properties. Driver capabilities (i.e., capacity) are determined by many factors including biological factors, knowledge, skills, and allocation of resources. The "fit" between driver demands and capabilities contributes to the perceived difficulty of the tasks.

Safe or successful performance is therefore contingent upon the ability of drivers to recognize the relationship between demands of the driving task and their own capabilities. When demands of the driving task exceed a driver's capabilities, the well-calibrated driver will recognize this difficulty and take measures to bring the two into alignment. For example, by reducing speed, a driver may ease the time pressure and challenges of vehicle control associated with a curvy stretch of road. Drivers that are not well-calibrated might fail to take protective countermeasures thereby placing themselves at risk (e.g., Deery, 1999; Spolander, 1983).

[1] We do note that, here and elsewhere, these studies often define or contrast between overconfidence, overestimates or over-placement in different ways. A fair and cogent treatment of these issues, however, is beyond the intended scope of the current effort.

While these models (Fuller, 2005; de Craen, 2010; Mitsopoulos-Rubens, 2010) share common features with respect to demand regulation, they differ in terms of the underlying constructs and driving factors. For example, de Craen (2010) views calibration as a deliberative process that is driven by motivational factors. In contrast, Mitsopoulos-Rubens considers calibration as a metacognitive process, one that is grounded in information processing. Furthermore, Mitsopoulos-Rubens (2010) underscores the importance of feedback in influencing knowledge of task demands and knowledge of capabilities at both a local, specific level and at a more global level. That is, feedback can update and refine the immediate processing of current information and action selection (local level), but can also inform more distal and general appraisals of self and competence (global level), as described earlier.

Since models of calibration require appraisals of task demands as well as driver capabilities and comparisons between the two, it follows that errors in calibration can arise from over- or underestimation in both or either of these constructs and/or from failures in the comparator mechanism. Obviously, underestimating task demands and overestimating capabilities carries the strongest implications for safety. A clear and accurate appraisal of task demands calls for an accurate perception of the state of the world and current performance whereas the realistic appraisal of capabilities relies on accurate knowledge of self, generally-speaking, as well as in situ.

While these models of demand regulation offer a very useful way of conceptualizing appraisals of situational factors and perceived capabilities as well as the resulting behavioral modification, they tend to offer a rather high level view of the relevant constructs. The mechanisms, although implied in many of the models, are not necessarily explicitly stated and the models are sometimes silent with respect to the many factors that can impact self and situational appraisals (and this is not to suggest that the authors of these models are unaware of these factors). In the following section, we propose a framework for calibration in the driving context that expands on these earlier models of momentary demand regulation by integrating aspects of human information processing (e.g., Wickens and Hollands, 2000) and lens models for information selection and utilization (Brunswik, 1955; Hammond, 1955). We believe that the inclusion of these other components offers a worthwhile expansion of the conceptual space and allows for a more pointed exploration of calibration and its role in road safety. For example, the inclusion of the information processing model allows one to drill deeper into cognitive processes that encompass perception and action in the world, highlighting the important role of attention in regulating the flow of information. The emphasis on attention in the model also allows for a more expansive discussion of internal and external factors that can impact its availability. Although prominent in other disciplines, the lens model has received little treatment in the driving domain even though it offers an effective conceptualization and quantification of information processing.

3. Framework development

It is because we are limited in our capacity, that we cannot process all available information and the information we do process is selective and subject to top-down as well as bottom-up influences. Many models of selective attention and eye scanning describe how we select specific visual information (e.g., Senders, 1983; Moray, 1986; Itti and Koch, 2000; Bundesen, 1990) and, oftentimes, we engage in shortcuts or heuristics in order to preserve cognitive resources (e.g., Kahneman and Tversky, 1972; Tversky and Kahneman, 1974; Slovic et al., 1977, 1979; Todd and Gigerenzer, 2003). It follows that a useful framework should incorporate a mechanism through which different information cues are selectively attended to, and weighed, in rendering appraisals or judgments. Lens models offer one such approach.

3.1. Lens model: selection and utilization of information

The lens model is a useful technique for exploring the nature and the quality of an observer's assessment of the world compared to the actual state of the world. The original lens model characterized observer perception (Brunswik, 1955); however, it was elaborated by Hammond (1955) in the context of judgment and decision making. At its core, the lens model describes how people use different information cues to make a judgment concerning the state of the world. The manner in which humans selectively weigh the information available to them (including, in some cases, a complete disregard of certain information) impacts the accuracy of their perception.

A simple lens model is shown in Fig. 1. The state of the world (SOW) is characterized as the environmental criterion variable (Y_E) and indices or proximal cues (X_1,X_2, . . . , X_n) provide information regarding the environmental criterion. The strength of the correlation between the cues and the criterion variable is a reflection of the ecological validity of each cue (w_{E1}, w_{E2}, . . . , w_{En}). On the right hand side of the figure is the subjective judgment (Y_J), where the observer weighs the various information cues (i.e., cue utilization policy) and renders a judgment or assessment of SOW. In practice, human judges might not weigh the information cues according to their informativeness of the true SOW and therefore the subjective judgment is only an approximation of the true SOW. The observer could also fall into a trap of overemphasizing or disregarding particular cues, relative to their utility (e.g., $w_{S1} \neq w_{E1}$). This could lead to very disparate estimates of the world. That is, as suggested by Dunning et al. (2004), errors in calibration can arise from incomplete processing or inappropriate use/weighing, as well as outright neglect, of available information in rendering a judgment. In this sense, the cue utilization strategy or policy employed by the observer can be likened to various heuristics used in theories of human judgment and decision making (e.g., Simon, 1956; Tversky and Kahneman, 1974; Ölander, 1975; Todd and Gigerenzer, 2003). These heuristics offer potential shortcuts for preserving cognitive resources, representing a more deliberate strategy for limiting the information that is processed.

The lens model can be readily quantified and distilled into a number of important metrics concerning the accuracy of an

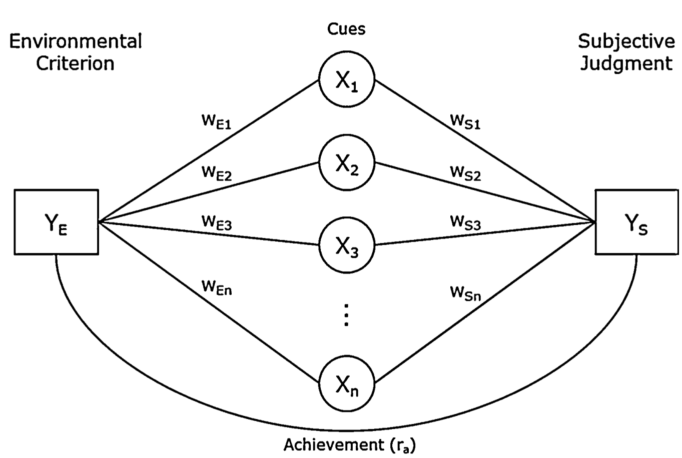

Fig. 1. Schematic representation of the lens model. Shown are various information cues (X_n), the weighting of these cues (w) both by the observer (S) and in relation to the environmental criterion (E). Achievement (r_a) reflects the degree of correspondence between the judgment and the environmental criterion (i.e., the accuracy of the judgment). After Goldstein (2006).

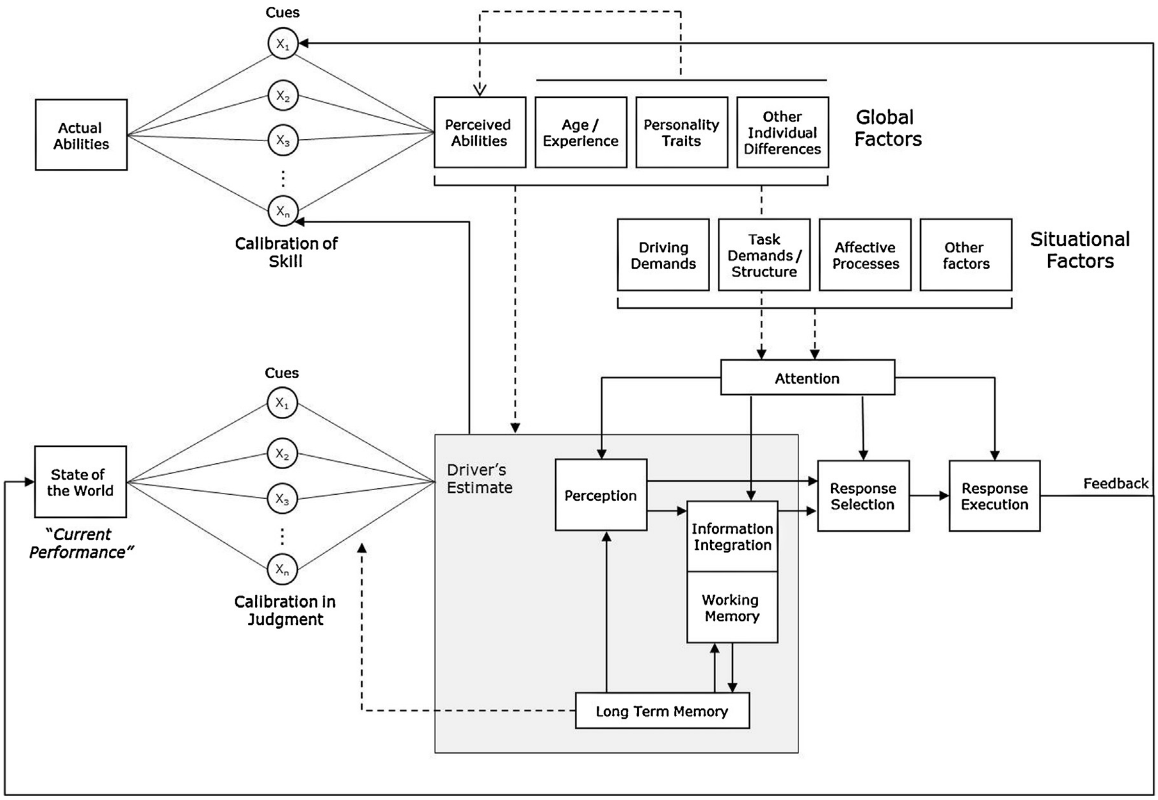

Fig. 2. Framework for examining calibration in driving.

observer's judgment as well as some of their underlying biases and has been applied in a wide range of domains, including the assessment of diagnostic skills of physicians, weather forecasters, auditors, pilots, air traffic controllers, and many others (e.g., Murphy, 1988; Cooksey, 1996; Stewart, 1990; Kirlik, 2006; Hammond and Stewart, 2001).

3.2. Driver calibration framework

Our conceptual framework is shown in Fig. 2. The lower half of the figure captures the momentary appraisal of the world as well as the driver's subsequent response and action. On the lower left, a lens model characterizes the manner in which drivers assess SOW—in this case expressed as "current performance", but could also be expressed in terms of task demands or some other environmental criterion. On the right side of the lens model, the driver uses the information cues (X_1, X_2, ..., X_n) to make a subjective estimate of the state of the world. Elaborating on the simple lens model shown in Fig. 1, this process is built into a model of information processing (shown on the lower right; after Wickens and Hollands, 2000) that invokes perception and various aspects of cognition (memory and information integration) in rendering a judgment.[2] It follows that the driver's estimate of SOW is determined by the information that the driver selectively attends to and by the manner in which the information is weighed and integrated (drawing upon information stored in long term memory). It is important to note that top-down influences are not the sole determinants of information selection and processing; this can also be impacted by bottom-up factors such as the salience of information (e.g., Itti and Koch, 2000; Wickens et al., 2003). Also, drivers' estimates can be influenced by their perceived abilities and skills. For example, drivers who believe they are highly competent might have a different cue utilization policy (i.e., weighing of information cues) that serves to reinforce this belief. Here, calibration in judgment is defined as the correspondence between an observer's subjective estimate of SOW and the objective measurement of it.

Following the appraisal of SOW, drivers will select an appropriate response and execute it—an outcome that will affect the state of the world (via feedback). Thus, the process of sampling, judging, and acting upon the world is a closed-loop process. Response selection, in this model, assumes some of the function of the comparator component in earlier demand regulation models (although we note that the comparator function in these other models more often involves a direct comparison of perceived capabilities and perceived task demands; e.g., Davidse et al., 2010; Mitsopoulos-Rubens, 2010). That is, the response selection in the current framework is a reflection of the momentary needs based on the driver's evaluation, but is also influenced by many global factors, including the driver's perceived abilities.

The individual stages that constitute the information processing model are all serviced in varying degrees by attention. That is, perceiving, integrating, deciding and acting are all effortful processes—though not necessarily conscious, deliberative processes—and attention is a limited resource (e.g., Kahneman, 1973; Wickens, 2002). Global factors, such as age, experience, and domain expertise, can impact the availability of attentional resources and, because of potential limitations, we simply might not have the capacity to select, process, and integrate all the

[2] We adopt the information processing perspective because it offers a useful way of conceptualizing the role of attention and other underlying cognitive and response processes. Other models, such as Neisser's (1976) perceptual cycle model, could also be useful expressions of the interplay between knowledge, directed perceptual exploration, and the actual environment.

available, or even the most relevant, information. That is, our cue utilization policy is influenced by the amount of information that can be (or is) taken in at a given moment.[3] Situational factors can further restrict the availability of resources, by way of capacity (e.g., fatigue) or competition (e.g., from the demand or structure of other concurrent tasks). For example, as workload increases, as in the case of elevated driving demands or competition from multiple tasks, there may be fewer attentional resources available so drivers might change their utilization policy such that they only take into account the information cues that they consider the most relevant, while those viewed as less informative will be neglected. By this process, they can render a judgment that might be sufficient, although not optimal, while preserving attentional resources for other tasks (e.g., a process sometimes referred to as satisficing; Simon, 1956; Ölander, 1975). However, in other situations, a shortage of attentional resources can lead to misleading or false appraisals of SOW with potentially dangerous repercussions. In their meta-synthesis, Zell and Krizan (2014) found that self-evaluations were worse for tasks that were less familiar or more complex—those imposing greater demands on attention. Additionally, and following from Lee and See (2004), affective (emotional) processes that are context-driven can impact the manner in which information is selected and processed.

A second lens model in the upper left of Fig. 2 shows various information cues that guide the more general perception of skill or ability. Similar to the appraisal of SOW, drivers will selectively weigh and utilize different bits of information in making such self-evaluations. The degree of correspondence between their perceived abilities and their actual abilities can be referred to as the calibration of skill.[4] As noted by Dunning et al. (2004), disparities in self-appraisal can be due to incomplete or inappropriate processing of information or by neglect of information (here a reflection of the cue utilization policy). Importantly, and as shown in Fig. 2, the driver's estimates of the world and feedback generated by the response execution stage can themselves become information cues that guide the perception of one's own ability. That is, local estimates of SOW (whether accurate or not) might serve to modify or reinforce perceived abilities over time or with repeated exposure. We also note that the perception of ability interacts with other global factors, such as age, experience, and personality traits (e.g., older drivers might deem themselves to be less capable than their younger counterparts). Individual factors, such as intelligence, achievement, locus of control, narcissism, and self-esteem, have also been found to be associated with the accuracy in self-appraisals (e.g., Mabe and West, 1982; Brown, 1986; Farwell and Wohlwend-Lloyd, 1998). Paradoxically, Kruger and Dunning (1999) showed that those that are less skilled tend to overestimate their skills more than those who are more skilled (later termed the Dunning–Kruger effect). They attribute this to differences in metacognitive ability: unskilled individuals simply do not have the capacity or tools to assess their own abilities effectively.

This framework conceptualizes calibration as it relates to momentary judgments of SOW and to appraisals of self-competence. The lens models offer a useful means of illustrating how drivers selectively attend to, and weigh, relevant information (and the computational aspect of the lens model, described elsewhere, can provide a useful means of quantifying the relationships; e.g., Cooksey, 1996). The stages of the information processing model reflect the agency of drivers in the world and also highlight the importance of attention in making local judgments. Finally, the important roles of both situational factors (the current context) as well as more global factors are elucidated (though we grant that the current lists of factors are illustrative, not exhaustive). Empirical evaluation of these and other relationships will allow for refinement of the framework.

One aspect of the model merits a final, yet important comment: the manner by which the environmental criterion (or state of the world) is defined or measured. The state of the world cannot be automatically known; it too must be measured and tools used to do so are imperfect and subject to misinterpretation and random noise (even objectively-defined ones). As suggested by Ackerman et al. (2002), some of the miscalibration between self-ratings and objective measures is due to a lack of specificity in the measures employed. Likewise, Dunning et al. (1989) suggested that some of the self-enhancements could be due in part to ambiguity in the characteristic dimension (i.e., what they are asked to appraise; e.g., trait could describe a wide variety of behaviors; echoed by Sundström (2008) in the driving context). For example, a rating of general driving skill could be interpreted in the context of crash history or history of moving violations and in terms of operational vehicle control, among many other facets. Thus, with respect to observed errors in calibration and in the absence of more specific guidance, people might select the most favorable criteria to make their assessment or they might simply consider the evidence that most readily comes to mind (availability heuristic, Tversky and Kahneman, 1974). As noted by Zell and Krizan (2014), the gaps between subjective and objective appraisals can be attenuated when the domain area or measure is more specific in nature.

4. Application areas

In this section, we briefly discuss the importance of calibration in a few topical areas and illustrate how the conceptual model can help situate current research findings and guide future efforts. The selection of these particular areas was driven by the experiences and expertise of the authors as well as the fact that they represent common, topical, and emerging areas in road safety.

4.1. Driver distraction

Today, drivers are able to perform more and more in-vehicle activities as new embedded and portable devices and technologies become increasingly available. Although new technologies may afford drivers enhanced productivity, there are obvious safety concerns to the extent that these devices and related activities detract attention from the driving task. There has been a proliferation of studies pertaining to distracted driving—one that largely parallels the explosive growth of the devices themselves. These studies typically examine the impact of distracted driving on various measures of performance or other outcomes (see Regan et al., 2009 for a review). Not surprisingly, when drivers divide their attention between multiple tasks, their performance on one task, or another, suffers (e.g., Caird et al., 2004; Horrey and Wickens, 2006).

Studies have found that drivers might not be well-calibrated to the effects of distracting activities on their performance. For example, Lesch and Hancock (2004) found that a priori ratings of confidence in dealing with distracting tasks were not related to actual performance while distracted. Moreover, the authors did not find any relationship between subjective ratings of performance or task demands and actual performance, suggesting that some drivers may not be aware of performance decrements while distracted. In an extension of this work, Horrey et al. (2008) directly compared drivers' subjective estimates of performance while distracted with

[3] This said, we do not suggest that resource limitations or attempts to conserve attentional resources are the sole determinants of one's cue utilization policy. The degree of processing can be influenced by other factors as well.

[4] Here we cede that there are many terms that could be used here interchangeably; nomenclature aside, the important point is that this rating involves a more general assessment of one's self (similar to many of those studies described in Section 2) as opposed to a momentary appraisal of the SOW.

actual distraction effects, along a number of measures of driving performance. The results from their study suggest that, for the most part, drivers' perceptions are not well-calibrated to the distracting effects of a cognitively demanding in-vehicle task. Similarly, Yang et al. (2012) found that drivers over-estimated the impact of in-vehicle tasks on driving performance yet under-estimated the impact on visual scanning; Crisler and Tyrrell (2011) found that drivers failed to recognize that their responses to objects and events were more impaired by secondary tasks than were responses in lane keeping tasks.

In the context of the framework in Fig. 2, increased task demands in the form of concurrent activities affects the availability of attentional resources for the various stages of information processing. This can impact the effectiveness of response selection and execution (findings of degraded performance), but it can also impact the selection of relevant information, leading to poor calibration. For example, the presence of a secondary visual task might cause drivers to fail to process critical information because their visual attention is directed elsewhere. Poor calibration may also stem from a reduction in the availability of mental resources to effectively weigh and integrate information—that is, those resources that would normally be used in support of situation awareness (Wickens, 2001, 2002). Although not specific to performance judgments, a number of studies have documented declines in an operator's situation awareness in the face of heightened levels of workload (e.g., Endsley and Kaber, 1999). The driver's appraisal of SOW, in this case erroneous, can impact their decision-making and response selection. Drivers may engage in distracting activities simply because they do not realize that their performance is degraded while distracted (i.e., they are not sampling the appropriate information) or they may be overly confident in their skills and their ability to deal with distractions while behind the wheel.

4.2. In-vehicle automation and autonomous vehicles

Computerization of vehicles has begun to fundamentally change the role of the driver. Collision warning systems and driver assistance systems, such as adaptive cruise control, autonomous braking, and steering assist already relieve the driver of some of the demands of driving. In the near future, drivers may be able to fully delegate control to the vehicle for extended periods on certain types of roads. These increasingly automated vehicles do not simply relieve people of various driving tasks, but they change the nature of driving and introduce new tasks (Woods, 1996). These new tasks and the spare capacity that these systems can produce interact with the new information sources that can distract drivers.

Calibration can play an important role in drivers' understanding, agreement, and trust in automated systems as well as the utility of such technological innovations, particularly for systems that infer whether drivers are impaired (e.g., distracted) or for systems where drivers need to monitor system function and takeover when deemed necessary (e.g., level 3, NHTSA levels of automation). A discrepancy between the driver's and the system's estimates of the state of the world could result in operator-system conflicts, reduced trust in the system and, ultimately, system disuse (Lee and See 2004; Lee and Moray, 1994; Parasuraman and Riley, 1997).

In Fig. 3 a new component is added on the left hand side, representing some form of in-vehicle automation. Presumably, the automation will be monitoring the state of the world, concurrently with the "driver" and the information cues may or may not be the same as those used by drivers. The automated system, particularly those that afford active vehicle control, can also impact the state of the world directly. Moreover, the automation itself might become an information cue that drivers use in their own appraisal of SOW (whether by direct or indirect feedback). Some systems, such as those that employ learning algorithms might also be informed by the systems knowledge of the driver's abilities. In Fig. 3, the calibration of the automation is also relevant; as with human judges, the automation can also make erroneous estimates of SOW. Equally important—especially in the context of system trust—is the correspondence between the automation's and driver's appraisals of SOW. If they are compatible, the driver is more likely to use the system; however, if they are very divergent, issues and conflicts are certain to arise.

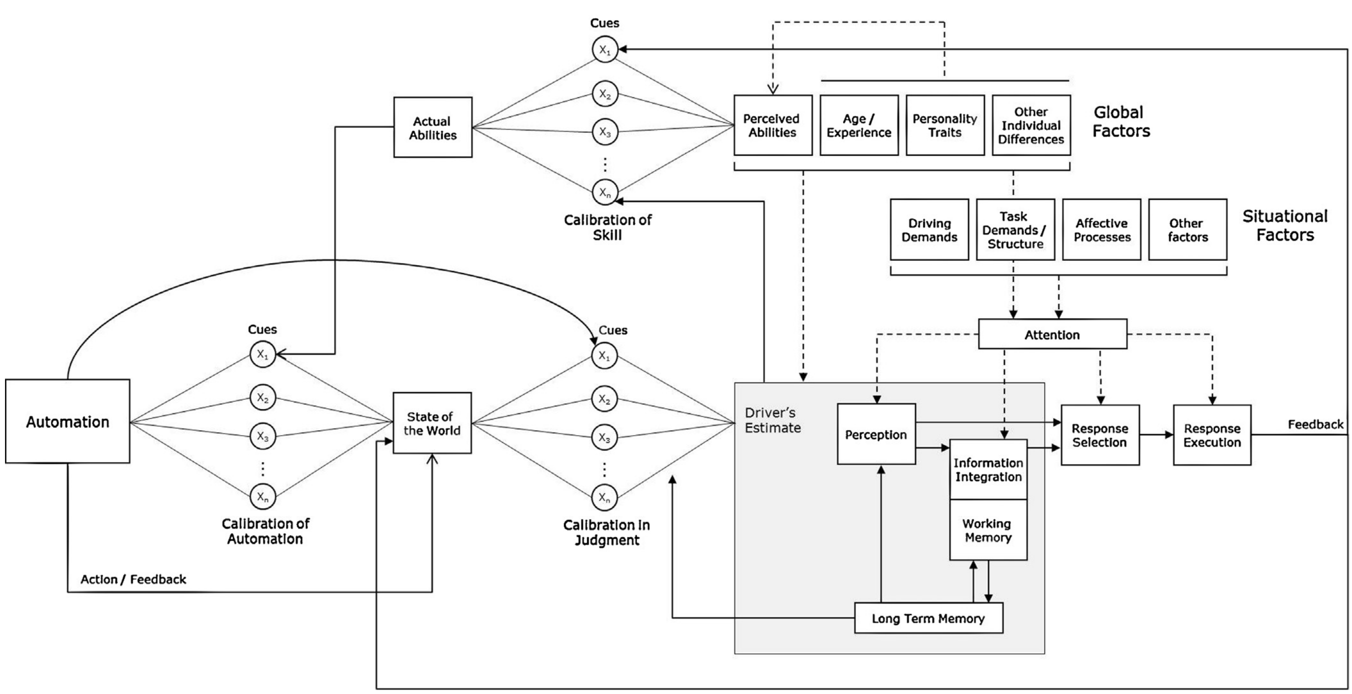

Fig. 3. Framework for examining calibration in driving, expanded to show the role and influence of automation (on the left hand side of the figure).

4.3. Inexperienced drivers

The notion of calibration is often implicated in discussions of inexperienced (typically younger) drivers. Indeed, some of the demand regulation models described earlier emerged from extensive work on inexperienced drivers and explorations into how errors in calibration can lead to inappropriate and unsafe road behaviors (cf., Mitsopoulos-Rubens, 2010; de Craen, 2010). Several studies have shown that novice drivers tend to overestimate their driving skills more so than experienced drivers (de Craen et al., 2011; Horswill et al., 2004) and others have found that such errors in calibration can lead to aggressive driving (Spolander, 1983). Some of the concerns over young and inexperienced drivers become more complex when one considers other potential confounding factors in their driving behaviors, such as their propensity towards potentially distracting technology (Lee, 2007).

Following from the framework shown in Fig. 2, it is clear that age and experience can impact a driver's perception of their own skills as well as the momentary appraisal of the state of the world. For example, an inexperienced driver might have unrealistic estimates of skill simply because they have not encountered sufficient driving situations with the appropriate degree of feedback (which feeds into their general appraisal of skill). Experience can thus influence how information regarding general skill is weighed and interpreted. As drivers gain more driving experience, they are also able to automatize some aspects of the driving task (e.g., Ranney, 1994), thus freeing up additional resources that might be used in support of the momentary assessment of the state of the world.

Given the importance of feedback in the framework—and in training, more generally—it follows that training can have important repercussions for both the development of actual skills as well as drivers' perceptions of these skills. Paradoxically, earlier work in driver training found that the training of different skills, such as skid control, actually led to worse safety outcomes (e.g., Elvik et al., 2009). Mayhew and Simpson (2002) cited overconfidence from the training course, coupled with poor retention of skills, as major determinants. Based in part on these outcomes, some individuals and agencies advocate training approaches that improve calibration skills such that drivers do not underestimate driving risks and overestimate their skills (e.g., Kuiken and Twisk, 2001; IRF, 2014). Gregersen (1996) nicely demonstrated that an "insight training" approach—one where drivers received direct experience and insight into their own limitations—could offset some of the undesirable patterns of overconfidence exhibited in a group that received skill-based training (see also Brookhuis et al., 2011; in the context of alcohol impairment). An insight approach (and other similar approaches) is consistent with the framework in Fig. 2, showing that feedback derived from actual experience or exposure can be used to update or reinforce general appraisals of skill and can also percolate to momentary appraisals of the world. In some ways, insight training can also provide more opportunities for pertinent feedback; in reality, there is often very little feedback for drivers to gauge how well they are doing—short of involvement in a crash or some other incident or traffic violation.

As a short sidebar, it is also important to note drivers at the other end of the spectrum. Some studies have shown that older drivers are also subject to similar errors or biases in calibration with respect to their skills and judgment (e.g., Freund et al., 2005; Horswill et al., 2004, 2013). However, many studies have shown that older drivers tend to self-regulate their behavior by avoiding certain types of driving situations such as nighttime or inclement weather or through active avoidance of certain intersections (e.g., Baldock et al., 2006; Molnar and Eby, 2008). Such behaviors could reflect a general awareness of their own deficits (i.e., good calibration). That said, older drivers remain overrepresented in motor vehicle crashes (e.g., Stutts et al., 2009), suggesting that these behavioral adaptations are not always effective. Although many factors can be implicated, the general cognitive and perceptual declines experienced with advancing age certainly impact the amount and degree to which information can be effectively processed—an outcome that could lead to important and safety-critical outcomes.

5. Conclusions

Humans often make inflated or erroneous estimates of their own ability or performance. Such errors in calibration are very important considerations in the context of road safety. The result can be poor decision making or risky behavior. Although the outcomes are often couched in different terminology or constructs, there is a vast literature that has documented the propensity for people to view themselves in favorable or optimistic light, irrespective of their actual performance, abilities or prospects (e.g., Zell and Krizan, 2014; Mabe and West, 1982). It is suggested that these biases or errors can result from incomplete or improper processing of available information (Dunning et al., 2004). In many cases, these appraisals represent more general estimates of self that could persist across time and different situations. Although not independent of these enduring traits, errors in estimates or judgments can also occur in the here and now; that is, errors in calibration can occur in the processing and determination of the current state of the world. Models of self-regulation nicely articulate the need for accurate appraisals of momentary demands, such that behavior can be adjusted accordingly (e.g., Fuller, 2005).

Building and elaborating on earlier models of calibration (demand regulation), we developed a framework that includes important aspects of information processing (Wickens and Hollands, 2000) and lens modeling (Brunswik, 1955). The former provides a stronger treatment of some of the underlying cognitive stages that are engaged in the selection and processing of relevant information as well as the process of decision-making (action selection) as well as execution. Moreover, it illustrates the important role of attention for all of these processes and highlights factors—both situational and global—that can impact its availability. The lens modeling component provides a conceptual (as well as mathematical) approach to examining how different information in the world can be selectively weighed and used by a human actor and allows one to contrast this cue utilization policy against the diagnosticity of this same information. With the proposed framework, the flow of information from selection, processing and integration to response execution not only impacts the state of the world (calling for the cyclic processing of information), but can also influence the more enduring self-appraisals.

Finally, we highlight a few sample areas where an understanding of calibration could shed some important insight into driver (or operator) decision-making and behavior as well as their performance and safety. These examples are merely illustrative and we believe that calibration could be an integral aspect to our understanding of human decision-making and behavior, not only in the driving context but also in many applied and safety-critical situations. While we acknowledge that the framework is not exhaustive or complete, we hope that it will be useful in articulating and guiding research efforts concerning the role of calibration in driving and other domains; in essence, helping account for discrepancies between perception, performance, and skill; and leading to approaches aimed at mitigating potentially adverse effects of miscalibration.

Acknowledgements

The authors are grateful to Yulan Liang, Amanda Young, Yusuke Yamani, Shannon Roberts, and Marvin Dainoff for helpful suggestions and guidance on earlier drafts of this manuscript and to Sarah Hong for her assistance.

References

Ackerman, P.L., Beier, M.E., Bowen, K.R., 2002. What we really know about our abilities and our knowledge. Pers. Individ. Differ. 33 (4), 587–605.
Baldock, M.R.J., Mathias, J.L., McLean, A.J., Berndt, A., 2006. Self-regulation of driving and its relationship to driving ability among older adults. Accid. Anal. Prev. 38, 1038–1045.
Baranski, J.V., Petrusic, W.M., 1994. The calibration and resolution of confidence in perceptual judgments. Percept. Psychophys. 55 (4), 412–428.
Baranski, J.V., Petrusic, W.M., 1995. On the calibration of knowledge and perception. Can. J. Exp. Psychol. 49 (3), 397–407.
Baumhart, R., 1968. Ethics in Business. Holt, Rinehart and Winston, New York.
Bol, L., Hacker, D.J., 2012. Calibration research: where do we go from here? Front. Psychol. 3, 1–6.
Brookhuis, K.A., De Waard, D., Steyvers, F.J.J.M., Bijsterveld, H., 2011. Let them experience a ride under the influence of alcohol; a successful intervention program? Accid. Anal. Prev. 43 (3), 906–910.
Brown, J.D., 1986. Evaluations of self and others: self-enhancement biases in social judgments. Social Cognit. 4 (4), 353–376.
Brown, I.D., Groeger, J.A., 1988. Risk perception and decision taking during the transition from novice to experienced driver status. Ergonomics 31, 585–598.
Brunswik, E., 1955. Representative design and probabilistic theory in a functional psychology. Psychol. Rev. 62, 193–217.
Bundesen, C., 1990. A theory of visual attention. Psychol. Rev. 97 (4), 523–547.
Caird, J.K., Scialfa, C.T., Ho, G., Smiley, A., 2004. Effects of Cellular Telephones on Driving Behaviour and Crash Risk: Results from a Meta-analysis. CAA Foundation for Traffic Safety, Edmonton, AB.
Cooksey, R.W., 1996. Judgment Analysis: Theory, Methods, and Applications. Academic Press, San Diego, CA.
Crisler, M., Tyrrell, R.A., 2011. Do drivers know that being distracted impairs their ability to respond to events more than their ability to maintain lane position? Transportation Research Board 90th Annual Meeting (No. 11-3745). Washington, DC, Transportation Research Board.
Davidse, R.J., Vlakveld, W.P., Doumen, M.J.A., de Craen, S., 2010. State Awareness, Risk Awareness of and Calibration by Road Users: A Literature Study. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.
de Craen, S., 2010. The X-Factor: A Longitudinal Study of Calibration in Young Novice Drivers. Unpublished Doctoral Dissertation. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.
de Craen, S., Twisk, D.A.M., Hagenzieker, M.P., Elffers, H., Brookhuis, K.A., 2011. Do young novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions. Accid. Anal. Prev. 43 (5), 1660–1665.
Deery, H.A., 1999. Hazard and risk perception among young novice drivers. J. Saf. Res. 30 (4), 225–236.
DeJoy, D.M., 1989. The optimism bias and traffic accident risk perception. Accid. Anal. Prev. 21, 333–340.
Dunning, D., Meyerowitz, J.A., Holzberg, A.D., 1989. Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. J. Pers. Social Psychol. 57 (6), 1082–1090.
Dunning, D., Heath, C., Suls, J.M., 2004. Flawed self-assessment. Psychol. Sci. Public Interest 5 (3), 69–106.
Elvik, R., Hoye, A., Vaa, T., Sorensen, M., 2009. The Handbook of Road Safety Measures, 2nd ed. Emerald, Bingley, UK.
Endsley, M.R., Kaber, D.B., 1999. Level of automation effects on performance: situation awareness and workload in a dynamic control task. Ergonomics 42, 462–492.
Farwell, L., Wohlwend-Lloyd, R., 1998. Narcissistic processes: optimistic expectations, favorable self-evaluations, and self-enhancing attributions. J. Pers. 66, 65–83.
Freund, B., Colgrove, L.A., Burke, B.L., McLeod, R., 2005. Self-rated driving performance among elderly drivers referred for driving evaluation. Accid. Anal. Prev. 37, 613–618.
Fuller, R., 2005. Towards a general theory of driver behaviour. Accid. Anal. Prev. 37, 461–472.
Goldstein, W.M., 2006. Introduction to Brunswikian theory and method. In: Kirlik, A. (Ed.), Adaptive Perspectives on Human-Technology Interaction. Oxford University Press, New York, pp. 10–24.
Gregersen, N.P., 1996. Young drivers' overestimation of their own skill–an experiment on the relation between training strategy and skill. Accid. Anal. Prev. 28, 243–250.
Griffin, D., Tversky, A., 1992. The weighing of evidence and the determinants of confidence. Cognit. Psychol. 24, 411–435.
Hammond, K., 1955. Probabilistic functionalism and the clinical method. Psychol. Rev. 62, 255–262.
Hammond, K.R., Stewart, T.R., 2001. The Essential Brunswik: Beginnings, Explications, Applications. Oxford University Press.
Hoorens, V., 1993. Self-enhancement and superiority biases in social comparison. Eur. Rev. Social Psychol. 4 (1), 113–139.
Horrey, W.J., Wickens, C.D., 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Hum. Factors 48 (1), 196–205.
Horrey, W.J., Lesch, M.F., Garabet, A., 2008. Assessing the awareness of performance decrements in distracted drivers. Accid. Anal. Prev. 40 (2), 675–682.
Horswill, M.S., Waylen, A.E., Tofield, M.I., 2004. Drivers' ratings of different components of their own driving skill: a greater illusion of superiority for skills that relate to accident involvement. J. Appl. Social Psychol. 34 (1), 177–195.
Horswill, M.S., Sullivan, K., Lurie-Beck, J.K., Smith, S., 2013. How realistic are older drivers' ratings of their driving ability? Accid. Anal. Prev. 50, 130–137.
International Road Federation, 2014. Training Drivers to Have the Insight to Avoid Emergency Situations, Not the Skills to Overcome Emergency Situations (Online Report). IRF, Alexandria, VI http://www.irfnews.org/wp-content/uploads/IRF-DBET-SC-Endorsement-Driver-Training-11-07-2013.pdf (accessed 09.09.14).
Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res. 40, 1489–1506.
Kahneman, D., 1973. Attention and Effort. Prentice-Hall, Englewood Cliffs, NJ.
Kahneman, D., Tversky, A., 1972. A judgment of representativeness Subjective probability. Cognit. Psychol. 3, 430–454.
Kirlik, A., 2006. Adaptive Perspectives on Human-Technology Interaction. Oxford University Press, New York.
Kruger, J., Dunning, D., 1999. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. J. Pers. Social Psychol. 77, 1121–1134.
Kuiken, M., Twisk, D., 2001. Safe Driving and the Training of Calibration. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.
Larwood, L., 1978. Swine flu: a field study of self-serving biases. J. Appl. Social Psychol. 8 (3), 283–289.
Larwood, L., Whittaker, W., 1977. Managerial myopia: self-serving biases in organizational planning. J. Appl. Psychol. 62 (2), 194–198.
Lee, J.D., 2007. Technology and teen drivers. J. Saf. Res. 38 (2), 203–213.
Lee, J.D., Moray, N., 1994. Trust, self-confidence: and operators' adaptation to automation. Int. J. Hum. Comput. Stud. 40, 153–184.
Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Hum. Factors 46 (1), 50–80.
Lesch, M.F., Hancock, P.A., 2004. Driving performance during concurrent cell phone use: are drivers aware of their performance decrements? Accid. Anal. Prev. 36 (3), 471–480.
Lichtenstein, S., Fischhoff, B., 1977. Do those who know more also know more about how much they know? Organ. Behav. Hum. Perform. 20 (2), 159–183.
Lichtenstein, S., Fischhoff, B., Phillips, L.D., 1982. Calibration of probabilities: the state of the art to 1980. In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, pp. 306–334.
Mabe, P.A., West, S.G., 1982. Validity of self-evaluation of ability: a review and meta-analysis. J. Appl. Psychol. 67 (3), 280–296.
Mayhew, D.R., Simpson, H.M., 2002. The safety value of driver education and training. Inj. Prev. 8 (Suppl. 2), ii3–ii8.
McKenna, F.P., Stanier, R.A., Lewis, C., 1991. Factors underlying illusory self-assessment of driving skill in males and females. Accid. Anal. Prev. 23 (1), 45–52.
Mitsopoulos-Rubens, P., 2010. Calibration Ability and the Young Novice Driver. Unpublished Dissertation. Monash University, Melbourne, Australia.
Molnar, L.J., Eby, D.W., 2008. The relationship between self-regulation and driving-related abilities in older drivers: an exploratory study. Traffic Inj. Prev. 9, 314–319.
Moray, N., 1986. Monitoring behavior and supervisory control. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Human Performance, vol. 2. John Wiley & Sons, New York, pp. 40.1–40.51.
Murphy, A.H., 1973. A new vector partition of the probability score. J. Appl. Meteorol. 12 (4), 595–600.
Murphy, A.H., 1988. Skill scores based on the mean square error and their relationship to the correlation coefficient. Mon. Weather Rev. 166, 2417–2424.
Neisser, U., 1976. Cognition and Reality: Principles and Implications of Cognitive Psychology. Freeman, San Francisco, CA.
Ölander, F., 1975. Search behavior in non-simultaneous choice-situations: satisficing or maximising? In: Wendt, D., Vlek, C. (Eds.), Utility, Probability, and Human Decision Making. D. Reidel, Dordrecht-Holland, pp. 297–320.
Parasuraman, R., Riley, V., 1997. Humans and automation: use misuse, disuse, abuse. Hum. Factors 39 (2), 230–253.
Ranney, T.A., 1994. Models of driving behavior: a review of their evolution. Accid. Anal. Prev. 26 (6), 733–750.
Regan, M.A., Lee, J.D., Young, K.L., 2009. Driver Distraction: Theory, Effects, and Mitigation. CRC Press, Boca Raton, FL.
Senders, J.W., 1983. Visual Sampling Processes. Erlbaum, Hillsdale, NJ.
Sharot, T., 2011. The optimism bias. Curr. Biol. 21 (23), R941–R945.
Simon, H.A., 1956. Rational choice and the structure of the environment. Psychol. Rev. 63, 129–138.
Simons, D.J., 2013. Unskilled and optimistic: overconfident predictions despite calibrated knowledge of relative skill. Psychon. Bull. Rev. 20 (3), 601–607.
Slovic, P., Fischhoff, B., Lichtenstein, S., 1977. Behavioral decision theory. Ann. Rev. Psychol. 28 (1), 1–39.
Slovic, P., Fischhoff, B., Lichtenstein, S., 1979. Rating the risks. Environ. Sci. Policy Sustain. Dev. 21 (3), 14–39.

Soll, J.B., 1996. Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. Organ. Behav. Hum. Decis. Processes 65 (2), 117–137.
Spolander, K., 1983. Accident Risks of Drivers: A Model Tested on Men and Women. National Swedish Road & Traffic Research Institute, Linkoeping, Sweden.
Stewart, T.R., 1990. A decomposition of the correlation coefficient and its use in analyzing forecasting skill. Weather Forecasting 5, 661–666.
Stone, E.R., Opel, R.B., 2000. Training to improve calibration and discrimination: the effects of performance and environmental feedback. Organ. Behav. Hum. Decis. Processes 83 (2), 282–309.
Stutts, J., Martell, C., Staplin, L., 2009. Identifying Behaviors and Situations Associated with Increased Crash Risk for Older Drivers (Report No. DOT HS 811 093). National Highway Traffic Safety Administration, Washington, DC.
Sundström, A., 2008. Self-assessment of driving skill–a review from a measurement perspective. Transp. Res. F Traff. Psychol. Behav. 11 (1), 1–9.
Svenson, O., 1981. Are we all less risky and more skillful than our fellow drivers? Acta Psychol. 47 (2), 143–148.
Todd, P.M., Gigerenzer, G., 2003. Bounding rationality to the world. J. Econ. Psychol. 24 (2), 143–165.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185 (4157), 1124–1131.
Weinstein, N.D., 1980. Unrealistic optimism about future life events. J. Pers. Social Psychol. 39 (5), 806–820.
Wickens, C.D., 2001. Workload and situation awareness. In: Hancock, P., Desmond, P. (Eds.), Stress, Workload, and Fatigue. Lawrence Erlbaum, Mahwah, NJ , pp. 443–454.
Wickens, C.D., 2002. Multiple resources and performance prediction. Theor. Issues Ergon. Sci. 3, 159–177.
Wickens, C.D., Hollands, J.G., 2000. Engineering Psychology and Human Performance, 3rd ed. Prentice Hall, Upper Saddle River, NJ.
Wickens, C.D., Goh, J., Helleburg, J., Horrey, W.J., Talleur, D.A., 2003. Attentional models of multi-task pilot performance using advanced display technology. Hum. Factors 45 (3), 360–380.
Woods, D.D., 1996. Decomposing automation: apparent simplicity, real complexity. In: Parasuraman, R., Mouloua, M. (Eds.), Automation and Human Performance: Theory and Applications. Erlbaum, Mahwah, NJ, pp. 3–17.
Yang, Y., McDonald, M., Reimer, B., Mehler, B., 2012. Are drivers aware of their behavior changes when using in-vehicle systems. Proceedings of 15th International IEEE Conference on Intelligent Transportation Systems (ITSC) 1515–1518.
Zell, E., Krizan, Z., 2014. Do people have insight into their abilities? A metasynthesis. Perspect. Psychol. Sci. 9 (2), 111–125.