ELSEVIER

Accident Analysis and Prevention 39 (2007) 1177–1185

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test

Thomas S.A. Wallis, Mark S. Horswill *

School of Psychology, University of Queensland, St. Lucia, Queensland 4072, Australia

Received 8 November 2006; received in revised form 12 January 2007; accepted 6 March 2007

Abstract

Drivers' hazard perception ability, as measured in video-based simulations, correlates with crash involvement, improves with experience and can be trained. We propose two alternative signal detection models that could describe individual differences in this skill. The first model states that novice drivers are poorer at discriminating more hazardous from less hazardous situations than experienced drivers. The second model proposes that novice drivers require a higher threshold of danger to be present before they notice a situation is hazardous or before they are willing to classify a situation as hazardous. We applied a technique involving fuzzy signal detection analysis to differentiate between these two models when comparing novice and experienced drivers, and trained and untrained drivers, in various video-based hazard perception measures. The data favored the second model.

Keywords: Driving behavior; Hazard perception skill; Training; Response bias; Sensitivity; Signal detection theory

1. Introduction

1.1. Hazard perception in driving

One of the few driver skills that has been found to correlate with crash involvement across a number of studies is drivers' hazard perception ability (Drummond, 2000; Hull and Christie, 1992; McKenna and Horswill, 1999; Pelz and Krupat, 1974; Quimby et al., 1986; Transport and Road Research Laboratory, 1979; see Horswill and McKenna, 2004, for a review). This is typically measured by presenting individuals with driver's perspective filmed footage of various traffic situations and asking them to indicate when they detect a potential traffic hazard, either by moving a handle (Pelz and Krupat, 1974; Watts and Quimby, 1979), using a touch screen (Hull and Christie, 1992), or pressing a response button (McKenna and Crick, 1991). An average reaction time to selected hazards is typically used as the measure of ability. This type of measure has been found to correlate with driving instructors' on-road ratings of individuals' hazard perception skill (Mills et al., 1998).

In addition, experienced drivers have been found to have faster reaction times to hazards than novice drivers (McKenna and Crick, 1991; Quimby and Watts, 1981; Sexton, 2000) and this is consistent with the finding that experienced drivers have a substantially lower crash risk than novices (Maycock and Lockwood, 1993); though note that the novice/experienced difference has not been found for all hazard perception tests (Chapman and Underwood, 1998; Crundall et al., 2003; Sagberg and Bjornskau, 2006).

Hazard perception has also been found to be amenable to training interventions. Both advanced police drivers and civilian drivers, who took part in an advanced driving course including hazard perception training, were significantly faster at perceiving hazards than experience-matched control groups (McKenna and Crick, 1991). Also, novice drivers trained in anticipating hazards (both on-road and video-based) have been found to have faster hazard perception scores than untrained controls (McGowan and Banbury, 2004; McKenna and Crick, 1994; McKenna et al., 2006; Mills et al., 1998).

In sum, hazard perception ability predicts crash risk, increases with driving experience, and is susceptible to training, and therefore is likely to have important implications for road safety. Indeed, hazard perception tests are now a compulsory component of driver licensing assessment in the UK and some states in Australia. However, little is known about the

* Corresponding author. Tel.: +61 7 3346 9520; fax: +61 7 3365 4466.
E-mail addresses: t.wallis@psy.uq.edu.au (T.S.A. Wallis), m.horswill@psy.uq.edu.au (M.S. Horswill).

0001-4575/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2007.03.003

precise nature of this ability and why some drivers outperform others.

1.2. *Why are experienced and trained drivers better than novices on hazard perception tests?*

While hazard perception test scores are typically based on reaction times, hazard perception is nonetheless a detection task and therefore issues related to signal detection theory are relevant. However, as noted by some authors (Groeger, 2000; Horswill and McKenna, 2004), these issues have yet to be directly empirically addressed.

Individual differences in hazard perception scores could reflect both sensitivity (drivers' ability to discriminate between hazardous and non-hazardous situations) and response bias (the threshold of perceived hazardousness above which drivers respond). Alternative hypothetical response patterns of novice and experienced drivers to hazardous and less hazardous situations are given in Fig. 1 for (a) a sensitivity model and (b) a response bias model.

1.2.1. *The sensitivity model (Fig. 1a)*

In this model, individual differences in hazard perception are due to novice drivers being less able than experienced drivers to distinguish the anticipatory cues in more hazardous situations from those in less hazardous situations. In this context, "hazardousness" refers to the likelihood that a collision or near collision with another road user will occur unless action is taken by the camera car driver. By the sensitivity model, novices are more likely to confuse anticipatory cues, leading them to misclassify the potential for a traffic conflict to occur in a scene. If this model is correct then we would expect differences in signal detection sensitivity measures (such as d') between the novice and experienced groups—experienced drivers will be more accurate at determining the level of hazard present in a scene.

How could the latency difference found between novices and experienced drivers map onto this model? In a typical hazardous incident, the level of hazard increases as the scene progresses. Initially there may be only scant evidence that a crash is possible, but this evidence increases as the scene develops, up until the point at which the crash would have occurred if no evasive action was taken by the driver. One possibility is that experienced drivers may accumulate this evidence faster than novices and hence respond earlier. In contrast, novices are slower at accumulating this evidence, and are more likely to misclassify hazards (they confuse non-hazards and hazards because they fail to interpret the initial evidence correctly).

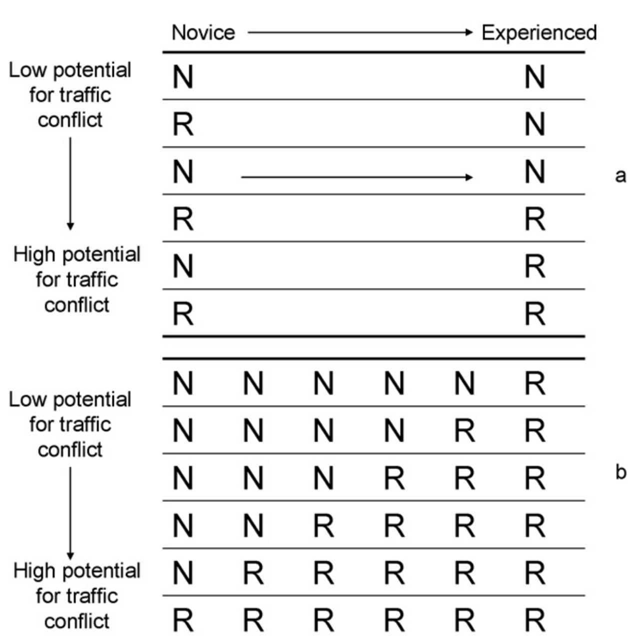

Fig. 1. Hypothetical response data for (a) pure sensitivity account and (b) pure response bias account. "N" represents no response, "R" represents a response. Columns are participants, rows are traffic scenes.

1.2.2. *The response bias model (Fig. 1b)*

In the response bias model, differences between novice and experienced drivers are not a result of misclassification of hazards by novices. Instead, novices only respond to the most hazardous situations that are presented while experienced drivers respond to a broader range of situations, which include less hazardous scenes. This pattern of results could reflect two alternative underlying mechanisms.

First, novices may respond more slowly to the less hazardous scenes simply because they only notice the hazards when the anticipatory cues become more obvious. If novices are proportionally slower to accumulate evidence of potential danger when that danger is less noticeable, then their reaction times will be slower. That is, differences in hazard perception are still a function of anticipation skill, but novices do not confuse less hazardous and more hazardous situations as proposed in the sensitivity model.

An alternative mechanism behind the response bias model is that novices can interpret the road situation just as accurately and as quickly as experienced drivers, but they are simply less willing to label situations as hazardous. That is, differences in hazard perception are due to differences in the definition of what subjective traffic conflict likelihood counts as a "hazard", rather than differences in anticipatory skill. Novice drivers may simply consider the same situation to be less risky than experienced drivers, who react in a more cautious manner. If the level of danger in a hazardous situation typically increases over time, then experienced drivers will reach their threshold of what is dangerous enough to be considered a hazard sooner than novices, generating the group reaction time differences observed.

1.3. *Available indirect evidence for experienced/novice differences*

There is, to date, no research evidence that rules out either one of the two models in Fig. 1. However, there is some indirect evidence that is relevant to the issue.

The tracking of drivers' eye movements has demonstrated that patterns of eye scanning differ between drivers of varying skill (Crundall et al., 2003; Crundall and Underwood, 1998; Mourant and Rockwell, 1972; Underwood et al., 2002, 2003). Expert police drivers showed a larger visual sampling rate and search spread than novice drivers, though they did not report more hazards in the scenes (Crundall et al., 2003). Experienced drivers have been shown to be flexible in adapting their eye-scanning patterns to different road types and traffic situations, whereas novice drivers were inflexible and showed longer fixation durations (Crundall and Underwood, 1998). If novices were seeing the same events at the same time as experienced drivers but were choosing not to respond, then no visual search differences between drivers of differing skill levels would be expected (Horswill and McKenna, 2004). These data indicate a difference in visual attention guidance ability between experienced drivers and novices, which suggests that they are not simply shifting their decision threshold. However, it could be argued that greater fixation durations of novices could be interpreted as a more conservative criterion-novices need to accumulate stronger evidence to exceed their threshold, and thus fixate for longer when scanning scenes.

If faster latencies in the hazard perception test were due to a more liberal response criterion (as suggested in Section 1.2.2) then more scenes would exceed the person's decision criterion and thus be labeled "hazardous". A participant responding with this strategy could be expected to respond more frequently on the test than others, as more scenes exceed their threshold. However, as noted above, Crundall et al. (2003) found that police drivers did not report more hazards in traffic scenes than novices. Also, McGowan and Banbury (2004) reported no association between hazard perception latency and total number of responses.

Additionally, if drivers who respond faster in hazard perception tests gain their advantage by simply being more cautious (and hence more willing to classify events as hazardous) then they would also be expected to rate traffic scenes as being more risky (Horswill and McKenna, 2004). Contrary to this, Farrand and McKenna (2001) found no association between hazard perception performance and ratings of risk. That is, participants who responded faster did not rate scenes as any more risky than those who responded slower. However, it could be argued that risk ratings represent an un-timed, out-of-task assessment of the hazardousness of a scene. Participants may agree on their ratings of the hazardousness of a scene irrespective of their experience level, but respond differently in task because of a different decision criterion. That is, even if experienced drivers are using a more liberal criterion to produce faster decisions, they may not use this criterion on the rating task.

1.4. The problem with using traditional signal detection theory to contrast the two proposed models

The traditional method for determining which of the two proposed models of hazard perception is true would be to conduct a signal detection analysis (Green and Swets, 1966; Macmillan, 2002; Macmillan and Creelman, 1991) in which measures of correct hits (when a participant correctly identifies a hazard) and false positives (when a participant incorrectly identifies a non-hazard as being a hazard) are entered into formulas to determine separate measures of sensitivity and response bias (Stanislaw and Todorov, 1999). However, there are a number of reasons why this approach is both conceptually inappropriate and practically difficult for hazard perception-like tasks.

In the hazard perception domain, there is no way to measure objectively whether a scene is "a hazard" or "not a hazard". Traffic environments can be considered to vary in their potential for hazard with context and over time. All traffic situations can be better conceptualized as potentially hazardous to some degree. Thus, traditional signal detection, which classifies stimuli into the binary categories of "signal" and "noise", is conceptually inappropriate for application to hazard perception, which lacks an objectively measurable assessment of a binary true state.

Hazard perception tests can be considered to share properties with vigilance tasks, which have also been argued to be unsuitable for traditional signal detection analysis (Todkill, 1990). Both tend to be prolonged, lack a discrete and uniform trial structure, and have targets are rare. For example, participants in vigilance tasks often have a zero or near zero false alarm rate (Swets and Kristofferson, 1970). Consequently, estimates of d' and β, the traditional indices of sensitivity and response bias, respectively, tend to be unstable and thus unusable (Todkill, 1990).

1.5. Fuzzy signal detection theory

One possible way around these problems is to use fuzzy signal detection theory (Parasuraman et al., 2000; Masalonis and Parasuraman, 2003), which is designed to allow signal detection theory analysis of fuzzy signals and fuzzy responses without imposing artificial dichotomies on the data. By using continuously scaled variables to represent the stimuli and the responses, fuzzy signal detection analysis avoids the loss of information and conceptual problems incurred by transforming a continuous dimension into the discrete levels of traditional signal detection theory.

Fuzzy signal detection theory re-partitions the outcomes of traditional signal detection theory across hit, miss, false alarm and correct rejection for each stimulus–response pairing, such that each can have partial membership in more than one outcome category (Parasuraman et al., 2000). On a given trial, each individual provides a response (between 0 "entirely no-like" and 1 "entirely yes-like") to a stimulus (between 0 "entirely non signal-like" and 1 "entirely signal-like"). Each trial is assigned membership in the four outcome categories using functions proposed by Parasuraman et al. (2000). These functions were designed to avoid illogical results for any inputs, and to reduce to traditional signal detection theory outcomes when given binary s and r inputs.

Take, for example, a response of 80% 'yes' to an event that is 60% signal-like (percentages are used here for ease of conceptualization). The event is somewhat signal-like so warrants a response (hit = 60%), but the individual over-responds so is assigned a proportion of false alarm (20%). There is also some membership in correct rejection (20%), as the response is not a definite yes and the stimulus has some representation in the noise category. There is 0% in the miss category as the response is stronger than the signal. The membership of the four outcomes for any one trial will always sum to 100%, representing the reconstituted traditional signal detection theory membership. If the response and signal values in the present example were rounded to 100% (to become like traditional binary signal detection), the outcome would be considered a hit.

Fuzzy hit rates, false alarm rates, miss rates and correct rejection rates are calculated by summing the membership of each outcome category across trials and dividing by the signal membership (for hit and miss) or not signal membership (for false alarm and correct rejection) across trials. As in traditional signal detection theory, hit rate and miss rate sum to 1, as do false alarm rate and correct rejection rate. These fuzzy rates are then used to calculate sensitivity and response bias as in traditional signal detection theory (Parasuraman et al., 2000).

1.6. The present experiment

We used fuzzy signal detection theory techniques to determine which of the two patterns of hazard perception responding illustrated in Fig. 1 was supported. We compared a group of novices who received no hazard perception training to a group of experienced drivers and a group of novices who did receive training. While training interventions have been found to be able to improve the hazard perception test latencies of novices (McKenna et al., 2006), it is possible that this improvement may reflect a different underlying mechanism than the novice to experienced driver difference. For example, while experienced drivers might gain their hazard perception advantage by developing better discriminatory ability, it is possible that trained novice drivers might gain their hazard perception advantage by shifting their decision criterion. That is, is the training giving novices the same skills as experienced drivers or is it merely enabling novices to resemble experienced drivers?

2. Method

2.1. Participants

Participants were 69 people (35 females) recruited using convenience samples (acquaintances of the first author), a first-year university participant pool (for course credit), and employment service advertising (participants were paid Aus$ 10, $n = 14$). No participants had previously taken part in a hazard perception test. Participants were required to hold a current Australian driver license (provisional or open). Novice drivers were defined as those who had been driving for 4 years or less on a provisional or open license, while experienced drivers were those who had been driving for 10 years or more and drove more than 8000 km/year on average. These definitions are in line with previous studies which have yielded skill differences between these groups (McKenna and Crick, 1991, 1994). Seventy-three people were tested in total, but 4 were excluded post hoc. One person was excluded for extreme under-responding, and three for extreme over-responding ($z > 3$), on the total number of responses across the test.

The experienced driver group did not receive training. Novice drivers were randomly assigned to a trained and an untrained group. This assignment resulted in three independent groups: trained novices ($n = 25$), untrained novices ($n = 27$), and untrained experienced ($n = 17$). The novice groups did not differ significantly on gender, age, kilometers driven per year on average, total number of years driving, accident involvement over the past 3 years, pure reaction time (with the overall mean substituted for missing values), advanced driving qualification, or prior hazard perception training (see Table 1 for demographic information). The experienced group was significantly older, $t(40) = 14.35$, $p < .001$ and $t(42) = 14.62$, $p < .001$, drove more kilometers per year on average, $t(40) = 3.84$, $p < .001$ and $t(42) = 4.51$, $p < .001$, and had held their license for a significantly longer time, $t(40) = 14.12$, $p < .001$ and $t(42) = 14.43$, $p < .001$, than trained and untrained novices, respectively.

2.2. Materials and apparatus

2.2.1. Ratings of traffic scenes by driving experts

As noted in Section 1.4, a problem with the domain of hazard perception is that there is no objectively quantifiable method of describing the level of risk in a traffic scene. In more subjective domains, the ratings of an authority in the field might be appropriate. For example, the ratings of professional driving instructors could be used as a benchmark for the level of risk present in a traffic situation (Crundall et al., 2003; McKenna and Crick, 1991; Mills et al., 1998), against which risk judgments by less experienced drivers can be compared.

Three experts were recruited: two civilian instructors and one police instructor. The first civilian expert had been driving for 37 years and instructing for 33, and drove 50,000 km/year on aver-

Table 1
Group demographics and driving information (S.D. in parentheses)

	Trained novices	Untrained novices	Experienced
Gender	12 F, 13 M	13 F, 14 M	10 F, 7 M
Age (years)	18.76 (1.30)	19.39 (1.20)	47.24 (9.85)
km/year	10,889 (7360)	8712 (7487)	25,806 (17,365)
Years driving	1.68 (1.12)	2.13 (1.16)	29.29 (9.74)
Accidents past 3 years (average)	0.72 (0.89)	0.81 (1.18)	0.47 (0.72)
Pure RT (ms)	251 (36)	250 (32)	256 (26)
Advanced training (yes *n*)	3	4	2
Prior haz perc training (yes *n*)	3	3	1

age. The second civilian instructor had been driving for 43 years and instructing for 10, and drove 86,000 km/year on average. Both civilian instructors held a Certificate 4 in Driver Training and Assessment. The third expert recruited for the study was a police driving instructor, with a Certificate 4 in Police Driver Training, who had been driving for 14 years and instructing for 3.5, and drove 16,000 km/year on average. Certificate 4 represents the highest qualification for driving instructors in the Australian system.

A range of genuine, unstaged hazardous traffic scenes were filmed from the driver's perspective. Video footage was reviewed by the first author and occlusions (blackouts) of scenes were created based on a judgment of large, moderate or little potential for traffic conflict after the occlusion point, with the aim of creating approximately equal numbers of scenes in these three categories. This resulted in 103 occluded scenes. Experts were instructed to estimate the potential for a traffic conflict to occur in the next few moments after occlusion, stipulated as less than 5 s after the occlusion point, if the camera car was to take no evasive action. A traffic conflict was defined as "a situation in which a collision or near collision with another road user (including stationery vehicles, cyclists, or pedestrians) would occur unless you take some type of evasive action (slowing, steering, etc.)". Responses were made using a 20 point scale (labeled from "no potential" to "unavoidable"). Note that experts were not asked to judge when to take action. We assessed the experts' subjective judgment of the strength of the anticipatory cues present in the scene before the occlusion, not their own propensity for delay in making a response.

The ratings of each expert to the 103 scenes were standardized against each individual's ratings of all the scenes. The scenes with the lowest standard deviations across the experts (indicating higher agreement between the three experts on the relative potential for hazard) were chosen to create a 20 scene hazard perception test sequence (with occlusions removed) and a 23 scene hazard rating sequence (with occlusions left in place). Independent groups *t*-tests revealed no significant differences between the two sequences on potential for hazard, $t(42) = .07$, $p = .94$, nor on expert agreement, $t(42) = .21$, $p = .83$. Inter-rater reliabilities for the two sequences were high, lying above $r = .86$ in all cases.

2.2.2. The hazard perception test

The hazard perception test video consisted of 20 traffic scenes viewed from a distance of 110 cm on a 68 cm television (same display used for all measures). Participants were instructed to imagine they were driving the camera car, and to press a response button as quickly as possible when they anticipated a potential traffic conflict occurring. The definition of "traffic conflict" was the same as that for experts (see Section 2.2.1). Note that we are not asking when evasive action would be taken (a measure of propensity for delay), but when the participant recognizes anticipatory cues that may lead to a traffic conflict. A reaction time to each of 27 selected events (some of the scenes had multiple hazards) was recorded (accurate to ±18 ms), and signal detection measures were calculated from 21 selected events which had been rated by experts.

2.2.3. Hazard rating task

The hazard rating task video consisted of 23 traffic scenes which were blacked out at selected points (see Section 2.2.1). Each scene was preceded by an onscreen number to aid identification. Participants were instructed to rate the potential for a traffic conflict to occur in the next few moments after the occlusion on the same 20-point rating scale used by the experts (see Section 2.2.1). Participants marked their ratings for each scene on a response sheet.

2.2.4. Simple reaction time test

Participants' simple reaction time was assessed using a display consisting of a dark blue dot (3.5 cm screen diameter) displayed for 1 s on a light blue background, for 16 trials. Participants were required to press the response button as quickly as possible when the dark blue dot appeared on the screen.

2.2.5. Hazard perception training

The hazard perception training video was developed by Raikos (2003) based on a technique used by McKenna et al. (2006) and consisted of 17 min of hazardous traffic scenes with an instructional commentary by a professional driving instructor (not one of the experts used in the current study). This commentary was used to provide an instructional framework for participants, in terms of recognizing cues for potential hazards, anticipating potentially hazardous situations, and possible courses of action to avoid hazards. Some examples of commentary include "checking the intersections, looking for cars pulling out" and "person running across road up ahead, other person may follow". The untrained novices and the experienced drivers viewed a control video, which consisted of the same 17 min of traffic footage as the training video but without the instructional commentary. Raikos (2003) found that drivers who watched the video with commentary reacted significantly (949 ms) faster on average than drivers who watched the video with no training commentary.

2.3. Procedure

A questionnaire was used to collect demographic information and relevant driving information. Participants assigned to the untrained condition (including all experienced drivers) were given written instructions to watch the control video and imagine they were driving the camera car. Those assigned to the trained condition were instructed that they would be trained to anticipate potential traffic hazards by watching a 17-min video (the training video). They were asked to imagine they were driving the camera car, to attend to the commentary, and to think about sources for potential hazards in a given situation by asking themselves what might be reasonably expected to happen. Following the video, participants completed the simple reaction time test, the hazard perception test, and finally the hazard rating task. Note that for the hazard perception test, a conservative instruction manipulation was used for half of the trained novices and half of the untrained novices, in which they were instructed that they would be penalized if they pressed when there was no hazard. This had no effect on the results (all *F*s were ns), so has been excluded from the analysis. The order of presentation of the hazard perception test and the hazard rating task could not be counterbalanced because of this manipulation.

3. Results

3.1. Collation of variables

We designed the full contingent of 27 events on the hazard perception test to represent the full range of hazardousness (according to the experts' ratings). This was to enable signal detection measurement by deliberately including some scenes that some participants did not respond to. Reaction time measures could not be based on the full contingent of scenes because some of the scenes elicited no responses (as we intended). Events were only included in the calculation of reaction time if 90% or more of participants in each group responded. This criterion was to avoid the average reaction time of groups who responded less across the test reflecting more mean substitution than other groups. This resulted in 15 events being excluded from the analysis, leaving 12 events. Missing values in these events were replaced by mean reaction times for each. The mean latency of these 12 events was used as the measure of reaction time performance on the hazard perception test. Except where noted, other variables in the study refer to all 27 scenes.

Signal detection analysis of the hazard perception test was performed by assigning a binary yes/no value to the reaction time response based on the point at which the hazards were occluded when they were shown to the expert raters, and comparing this to experts' subjective ratings of hazardousness. Expert ratings were collected for 21 of the 27 events in the hazard perception test, hence these analyses are based on these 21 events. Responding before the occlusion point shown to experts was classified as a "yes" and responding after or not at all was classified as a "no". Signal detection measurements in the hazard rating task were performed using participants' raw ratings transformed into a range between 0 and 1, with 0 representing a no response (or no potential for hazard), and 1 representing a yes response (or unavoidable hazard). Formulae derived from Parasuraman et al. (2000) were used to calculate measures of signal detection. Fuzzy hit rates and false alarm rates of 0 or 1 were adjusted to .001 and .999, respectively, to allow calculation of signal detection variables.

3.2. Latency differences and total responses on the hazard perception test

Experienced drivers responded significantly faster on average to the hazard perception test than untrained novices, $t(42) = 3.88$, $p < .001$, and trained novices responded faster than untrained novices, $t(50) = 3.08$, $p < .01$ (see Fig. 2). The 12 scenes used to calculate latency were found to be acceptably internally consistent, $\alpha = .74$.

The more often participants pressed the response button over all test scenes, the faster their latency for the key scenes, $r = -.58$, $n = 69$, $p < .001$. Both the trained novice group, $t(33.04) = 2.83$, $p < .01$, and the experienced group, $t(18.03) = 4.75$, $p < .001$, responded significantly more times during the hazard perception test video than the untrained novice group (Fig. 3).

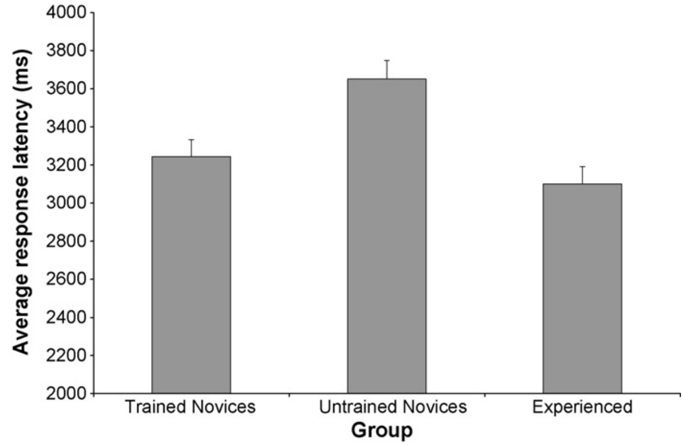

Fig. 2. Mean response latency (+S.E.M.) in the hazard perception test (12 events) by group.

3.3. The hazard rating task

The average rating across the 23 scenes in the hazard rating task ($M = 7.53$, S.D. = 2.49, skewness $z = 0.06$, kurtosis $z = 0.02$) was not subject to ceiling or floor effects, which can be a problem for this type of measure. Mean hazard rating was not significantly correlated with mean hazard perception test latency, $r = -.07$, $n = 69$, $p = .58$. The three groups did not differ significantly in their hazard ratings (trained-untrained $t(50) = 0.18$, ns, experienced-untrained $t(42) = 1.14$, ns, trained-experienced $t(40) = 0.88$, ns).

3.4. Fuzzy signal detection analysis

Fuzzy signal detection analyses were performed on the hazard perception test and the hazard rating task using a binary and continuous classification of response respectively. Sensitivity was indexed using d', where larger values indicate greater

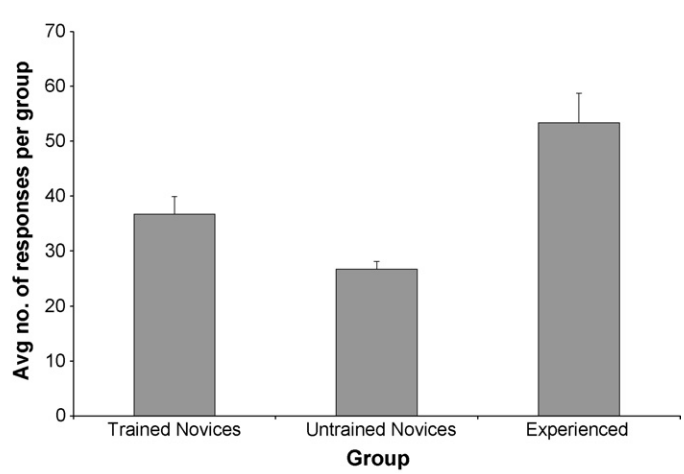

Fig. 3. Average number of responses (+S.E.M.) over the hazard perception test by group.

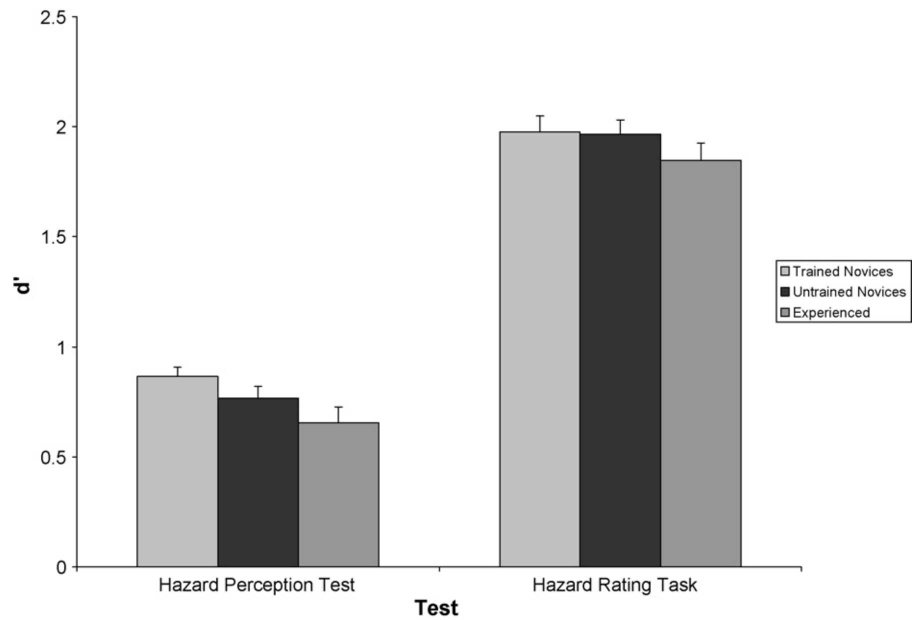

Fig. 4. Sensitivity measures (+S.E.M.) for the hazard perception test and the hazard rating task by group.

separation of signal and noise, and response bias was measured using *c*, where positive values indicate a conservative bias and negative values indicate a liberal bias (Stanislaw and Todorov, 1999). The parameter *c* was used as an estimate of response bias as it has been shown to be more appropriate for vigilance-like tasks (See et al., 1997).

Experienced drivers did not exhibit greater sensitivity than novices in the hazard perception test, $t(42) = 1.27$, ns, or in the hazard rating task, $t(42) = 1.16$, ns (Fig. 4). Similarly, the trained and untrained drivers did not differ in sensitivity in either task, $t(50) = 1.42$, ns and $t(50) = 0.11$, ns. Sensitivity on the hazard perception test and the hazard rating task was uncorrelated with latency on the hazard perception test for all groups, all rs < .25, ns.

The response bias measures reflected the latency differences between the groups better (Fig. 5). The untrained novice group was significantly more conservative than both the trained novice group, $t(50) = 3.14$, $p < .01$, and the experienced group, $t(42) = 3.83$, $p < .001$, on the hazard perception test. However, these differences did not carry over to the hazard rating task, $t(50) = 0.22$, ns, and $t(42) = 1.18$, ns. Response bias on the hazard perception test was significantly correlated with latency, such that more liberal responses were associated with faster latencies, for trained novices, $r = .85$, $n = 25$, $p < .001$, untrained novices,

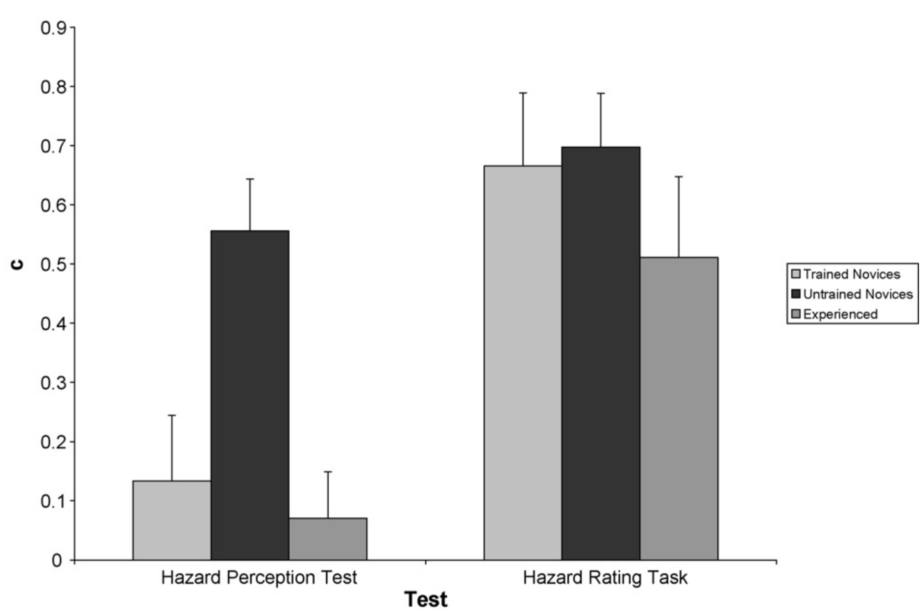

Fig. 5. Response bias measures (+S.E.M.) for the hazard perception test and the hazard rating task by group.

$r = .67$, $n = 27$, $p < .001$, and experienced drivers, $r = .80$, $n = 17$, $p < .001$. Response bias on the hazard rating task was uncorrelated with latency, all rs $< .09$, *ns*.

4. Discussion

The present study aimed to clarify which of the two competing models presented in Fig. 1 drives latency differences between experienced, trained and novice drivers on the hazard perception test. At least with respect to the present sample, the response bias model was supported (Fig. 1b). A more liberal response criterion on the hazard perception task correlated with latencies, consistent with the idea that a liberal response bias led to faster responses. Trained and experienced groups responded more liberally and faster than untrained novices. These results suggest that a pure sensitivity account of hazard perception performance, as supported by some indirect evidence, cannot explain group differences.

A caveat for the present research is that the values of the signal detection measures are relative to experts' subjective ratings, and not an objective state of "signal" versus "noise". As such, this cannot strictly be considered a true signal detection theory analysis, which (as outlined in Section 1.4) is impossible to calculate in this context. However, relative group differences in our sensitivity and response bias measures still serve the purpose of distinguishing between the two alternative patterns of data presented in Fig. 1.

That experienced drivers and novices, and trained and untrained drivers, were no different in their ability to discriminate scenes on the potential for hazard fits with the lack of group rating differences found, and the absence of a relationship between hazard ratings and hazard perception reaction time. This corroborates the findings of Farrand and McKenna (2001). The groups agreed on the level of hazardousness present in the scene when rated un-timed and out-of task, but their in-task responding did not correspond to this. Hence, novice drivers are as good as experienced and trained drivers at judging the potential for traffic conflicts based on anticipatory cues. However, they are less likely and slower to respond to less hazardous scenes on the test.

This finding may be considered within the broader theoretical framework of behavioral differences between novices and experienced drivers, particularly in the area of risk-taking behavior. McKenna et al. (2006) demonstrated that improvements in hazard perception performance through anticipation training in novice drivers decreased risk-taking behavior. This suggests that the differential in-task response bias adopted by the groups may reflect less willingness to respond to less hazardous scenes by the untrained novices. However, whether this is the case or whether this is because they are slower to notice or process cues is a matter for future research.

The finding that more frequent responding was associated with faster latencies is contrary to the findings of McGowan and Banbury (2004). This may be due to the differing response modes of the tests: McGowan and Banbury used a procedure in which participants clicked on hazards with a computer mouse rather than a simple button press. The additional requirement to provide spatial definition of the event of interest may have caused participants to adopt a more conservative overall response strategy on their test.

The commentary training used was effective at producing performance improvements over the no-commentary control video, in line with previous research into hazard perception training (McGowan and Banbury, 2004; McKenna and Crick, 1991, 1994; McKenna et al., 2006; Mills et al., 1998). Training novices to anticipate environmental cues for potential hazards (for example, checking for cars pulling out of a slip lane) improved performance on the hazard perception test, apparently through the same response bias model that experienced drivers employ. Since hazard perception ability is related to crash risk (see Horswill and McKenna, 2004, for a review), these findings imply that novice training methods should focus on recognizing anticipatory cues, and encouraging anticipatory rather than delayed avoidance responding.

Acknowledgements

We would to thank Jay Frank for testing some of the participants, Mark Wetton and three anonymous reviewers for comments on a previous draft of this paper, Maria Raikos for preparing some of the stimuli including the training video, Joel Nielsen for providing the expert commentary on the training video, and Wayne Gane, Christine Filer, John Morton, and Sgt. Chris Edge for pilot testing and rating of hazard scenes.

References

Chapman, P., Underwood, G., 1998. Visual search of driving situations: danger and experience. Perception 27, 951–964.

Crundall, D., Chapman, P., Phelps, N., Underwood, G., 2003. Eye movements and hazard perception in police pursuit and emergency response driving. J. Exp. Psychol.: Appl. 9 (3), 163–174.

Crundall, D., Underwood, G., 1998. Effects of experience and processing demands on visual information acquisition in drivers. Ergonomics 41 (4), 448–458.

Drummond, A.E., 2000. Paradigm lost! Paradise gained? An Australian's perspective on the novice driver problem. In: Proceedings of the Novice Driver Conference (1st–2nd June), Bristol, Retrieved on 8th November 2006 from http://www.dft.gov.uk.

Farrand, P., McKenna, F.P., 2001. Risk perception in novice drivers: the relationship between questionnaire measures and response latency. Transportation Research Part F: Traffic Psychology & Behaviour 4 (3), 201–212.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, Oxford.

Groeger, J.A., 2000. Understanding Driving: Applying Cognitive Psychology to a Complex Everyday Task. Psychology Press Ltd., Hove.

Horswill, M.S., McKenna, F.P., 2004. Driver's hazard perception ability: situation awareness on the road. In: Banbury, S., Tremblay, S. (Eds.), A Cognitive Approach to Situation Awareness. Ashgate, Aldershot, UK.

Hull, M., Christie, R., 1992. Hazard perception test: the Geelong trial and future development. In: Paper presented at the Proceedings of the National Road Safety Seminar, Wellington, New Zealand.

Macmillan, N.A., 2002. Signal detection theory. In: Methodology in Experimental Psychology, In: Pashler, H., Wixted, J. (Eds.), Stevens' handbook of experimental psychology, vol. 4, third ed. John Wiley & Sons, Inc., New York, NY, pp. 43–90.

Macmillan, N.A., Creelman, C.D., 1991. Detection theory: A User's Guide. Cambridge University Press, New York, NY.

Masalonis, A.J., Parasuraman, R., 2003. Fuzzy signal detection theory: analysis of human and machine performance in air traffic control, and analytic considerations. Ergonomics 46 (11), 1045–1074.

Maycock, G., Lockwood, C.R., 1993. The accident liability of British car drivers. Transport Reviews 13 (3), 231–245.

McGowan, A., Banbury, S., 2004. Evaluating interruption-based techniques using embedded measures of driver anticipation. In: Banbury, S., Tremblay, S. (Eds.), A Cognitive Approach to Situation Awareness. Ashgate, Aldershot, UK.

McKenna, F.P., Crick, J.L., 1991. Hazard perception in drivers: a methodology for testing and training. Final Report. Transport Research Laboratory, Crowthorne, UK.

McKenna, F.P., Crick, J.L., 1994. Developments in hazard perception. Contractor Report 313. Transport Research Laboratory, Crowthorne, UK.

McKenna, F.P., Horswill, M.S., 1999. Hazard perception and its relevance for driver licensing. Journal of the International Association of Traffic and Safety Sciences 23 (1), 26–41.

McKenna, F.P., Horswill, M.S., Alexander, J.L., 2006. Does anticipation training affect drivers' risk-taking? J. Exp. Psychol.: Appl. 12 (1), 1–10.

Mills, K.L., Hall, R.D., McDonald, M., Rolls, G.W.P., 1998. The effects of hazard perception training on the development of novice driver skills. Final Report. Department for Transport, UK. Retrieved on 8th November 2006 from http://www.dft.gov.uk.

Mourant, R.R., Rockwell, T.H., 1972. Strategies of visual search by novice and experienced drivers. Human Factors 14 (4), 325–335.

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

Pelz, D.C., Krupat, E., 1974. Caution profile and driving record of undergraduate males. Accid. Anal. Prev. 6 (1), 45–58.

Quimby, A.R., Maycock, G., Carter, I.D., Dixon, R., Wall, J.G., 1986. Perceptual abilities of accident involved drivers. Final Report 27. Transport and Road Research Laboratory, Crowthorne, UK.

Quimby, A.R., Watts, G.R., 1981. Human factors and driving performance. Laboratory Report 1004. Transport and Road Research Laboratory, Crowthorne, UK.

Raikos, M.K., 2003. Improving speed in skilled anticipation: the case of hazard perception in driving. Unpublished Honors Thesis. University of Queensland, Brisbane.

Sagberg, F., Bjornskau, T., 2006. Hazard perception and driving experience among novice drivers. Accid. Anal. Prev. 38, 407–414.

See, J.E., Warm, J.S., Dember, W.N., Howe, S.R., 1997. Vigilance and signal detection theory: an empirical evaluation of five measures of response bias. Human Factors 39 (1), 14–29.

Sexton, B., 2000. Development of hazard perception testing. In: Proceedings of the Novice Driver Conference (1st–2nd June), Bristol, Retrieved on 8th November 2006 from http://www.dft.gov.uk.

Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods Instrum. Comput. 31 (1), 137–149.

Swets, J.A., Kristofferson, A.B., 1970. Attention. Annu. Rev. Psychol. 21, 339–366.

Todkill, A.L., 1990. Stimulus comparison strategies in vigilance like tasks. Unpublished Ph.D. Thesis. University of Queensland, Brisbane.

Transport Road Research Laboratory, 1979. A Hazard Perception Test for Drivers. Transport and Road Research Laboratory, Crowthorne, UK.

Underwood, G., Chapman, P., Bowden, K., Crundall, D., 2002. Visual search while driving: skill and awareness during inspection of the scene. Special issue: eye movements, visual attention and driving behaviour. Transportation Research Part F: Traffic Psychology & Behaviour 5 (2), 87–97.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., Crundall, D., 2003. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics 46 (6), 629–646.

Watts, G.R., Quimby, A.R., 1979. Design and validation of a driving simulator. Laboratory Report 907. Transport and Road Research Laboratory, Crowthorne, UK.