A predictive hazard perception paradigm differentiates driving experience cross-culturally

Phui Cheng Lim[a], Elizabeth Sheppard[b] & David Crundall[c]

[a] Corresponding author. Address: University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia. Tel: +603 8924 8350. E-mail address: khyx1lpn@nottingham.edu.my

[b] Address: University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia. E-mail address: elizabeth.sheppard@nottingham.edu.my

[c] Address: Nottingham Trent University, Nottingham NG1 4BU, UK. E-mail address: david.crundall@ntu.ac.uk

Abstract

Hazard perception (HP) tests are used in several developed countries as part of the driver licensing process, where they are believed to have improved road safety; however, relatively little HP research has been conducted in developing countries, which account for 80% of the world's road fatalities. Previous research suggests that drivers in these countries may be desensitized to hazardous road situations and thus have increased response latencies to hazards, creating validity issues with the typical HP reaction time paradigm. The present study compared Malaysian and UK drivers' HP skills when watching video clips filmed in both countries, using a predictive paradigm where hazard criterion could not affect performance. Clips filmed in the UK successfully differentiated experience in participants from both countries, however there was no such differentiation in the Malaysian set of videos. Malaysian drivers also predicted hazards less accurately overall, indicating that exposure to a greater number of hazards on Malaysian roads did not have a positive effect on participants' predictive hazard perception skill. Nonetheless the experiential discrimination noted in this predictive paradigm may provide a practical alternative for hazard perception testing in developing countries.

Keywords: hazard perception, driving experience, cross-cultural, what happens next, anticipation skill

A predictive hazard perception paradigm differentiates driving experience cross-culturally

1. Introduction

Among driving-specific skills, very few have been linked to accident likelihood with any reliability. Hazard perception (HP), or the ability to detect dangerous situations on the road, is one of these few. It is typically measured by showing drivers video clips of driving scenarios and asking them to respond as soon as they detect potential hazards. Shorter latencies are generally thought to indicate greater levels of skill, with various studies reporting that experienced drivers detect hazards faster than novices (Borowsky, Shinar, & Oron-Gilad, 2010; Horswill et al., 2008; Scialfa et al., 2011; Wetton et al., 2010; Wetton, Hill, & Horswill, 2011), and others finding a direct link between novice drivers' hazard perception latencies and their later accident involvement (Drummond, 2000; Wells, Tong, Sexton, Grayson, & Jones, 2008). As a result, hazard perception testing has been incorporated into the driver licensing process in several countries, with the UK and parts of Australia using the reaction time paradigm described above, and some researchers argue that evidence suggests this has improved road safety in the UK (Wells et al., 2008).

While the reaction time paradigm remains the de facto measure of hazard perception, its validity is not entirely without question. For instance, several studies have failed to find latency differences between experienced and novice drivers (Chapman & Underwood, 1998; Sagberg & Bjørnskau, 2006), with some suggesting that the efficacy of hazard perception testing depends largely upon the individual clips used. In one such example, Sagberg & Bjørnskau (2006) found that while their test as a whole did not differentiate between experienced and novice drivers, post-hoc examination indicated that certain clips did successfully differentiate experience, and the elements of complexity, surprise and anticipation appeared to be particularly important.

Thus far, the elements of a clip that successfully differentiates experience have mainly been suggested post-hoc (Borowsky, Oron-Gilad, Meir, & Parmet, 2011; Borowsky, Oron-Gilad, & Parmet, 2009; Sagberg & Bjørnskau, 2006), although others have investigated this more systematically (Crundall et al., 2012; Garay-Vega & Fisher, 2005). In a simulator study investigating different hazard types, Crundall et al. (2012) found that environmental prediction (EP) hazards, when the precursor and hazard were different but indirectly related, were better discriminators of experience than behavioral prediction (BP) hazards, when a hazard's precursor and the hazard itself were the same stimulus. These findings certainly lay the groundwork for a hazard typology, however, a substantial amount of work remains to be done to establish any reliable categorization, and the precise elements that elicit differences of experience have yet to be explored in depth.

While hazard perception has been explored and used in driver training in several developed countries, there has been little research done in low- and middle-income countries, where 80% of the world's road fatalities take place (Toroyan, 2013). For instance, in Malaysia, a middle-income country with a high percentage of car ownership, 2010 road fatalities were roughly seven times that of the UK when accounting for population differences (Toroyan, 2013). Lim, Sheppard, & Crundall (2013) conducted a cross-cultural hazard perception test, comparing latencies between Malaysian and UK drivers when watching video clips filmed in both countries. The study reported that Malaysian drivers had higher latencies than UK drivers across all clips but fixated hazards equally quickly, suggesting that their increased response times were due to an increased criterion for reporting hazards, rather than a difference in their ability to spot the visual cues that ultimately lead to hazard identification. The researchers concluded that reaction time in a Malaysian context was potentially unsuitable as a measure of hazard perception ability, and the possible criterion difference might stem from drivers in more dangerous environments becoming desensitized to hazards. As Malaysia's driving environment is fairly comparable to other developing countries in terms of on-road hazards and traffic fatalities (Toroyan, 2013), this raises the question of whether the traditional hazard perception test is suitable for many developing countries.

Given Lim et al. (2013)'s findings, the present study revisits a cross-cultural comparison between Malaysian and UK drivers, but employs a predictive paradigm unaffected by hazard criterion: the "What Happens Next?" test, previously found to differentiate experience among UK drivers (Jackson, Chapman, & Crundall, 2009). In the original study, drivers watched video clips containing hazards, but the clips were stopped and occluded immediately prior to hazard onset and drivers were asked to predict the events that might have occurred after this point. Jackson et al. (2009) found that experienced drivers predicted events more accurately than novices when all clip information was removed from the screen immediately following the occlusion point.

As the present study utilizes the same video clips as Lim et al. (2013)'s cross-cultural study, this provides an opportunity to draw comparisons between test paradigms. More importantly, it offers a measure of hazard perception that is unaffected by response criterion. Drivers are not asked to decide whether or not a hazard has occurred; they are merely asked to predict an event. The findings should help establish whether the cross-cultural differences seen in the previous study were entirely the result of a criterion difference, or also reflect differences in hazard perception skill. The present study also employs multiple choice questions unlike the original predictive paradigm, which used a free response format; this serves to establish a more viable version of the paradigm for large-scale testing, should it again differentiate experience.

We hypothesize that experienced drivers will outperform novices on all clips regardless of where they were filmed, however, as Lim et al. (2013) reported, we also expect this particular advantage to decrease when drivers view clips filmed in their non-home country.

2. Methods

2.1. Participants

Forty participants were recruited from the UK and 37 from Malaysia, all of whom held full, provisional or learner driving licenses from their respective countries and had normal or corrected-to-normal vision. Participants were split into two further sub-groups consisting of novice and experienced drivers, resulting in four groups in total: 19 UK novice drivers (mean age of 22.9 years and licensing time of 8.25 months, except for three learner drivers who had held their permit for 14, 90 and 48 months respectively), 21 UK experienced drivers (mean age of 23.3 years and licensing time of 54.9 months), 20 Malaysia novice drivers (mean age of 18.0 years and licensing time of 4.5 months) and 17 Malaysia experienced drivers (mean age of 22.5 years and licensing time of 55.8 months). Participants received either monetary compensation or course credit, where the latter was applicable.

2.2. Stimuli and apparatus

The original stimuli were the same videos used in Lim et al. (2013)'s cross-cultural study, consisting of 20 clips from Malaysia and 20 from the UK, each containing one hazardous event. Examples of these video clips can be seen in Figures 2 and 3. Each clip was edited to end immediately prior to hazard onset, while giving enough predictive information for a viewer to deduce or make an intelligent guess as to what would happen next (Jackson et al., 2009). The resulting clips ranged from 2.7 to 43.7 seconds in length. After each clip ended, a black screen was displayed for one second. Four numbered options then appeared on the screen describing four different possible scenarios that could have occurred after the occlusion point, one of which had actually taken place.

The four options for each video were determined via discussion between one Malaysian researcher and one UK researcher, both of whom held a full driving license in their respective countries. Each set of four options was different and unique to the video, and each option represented an event that could have feasibly taken place after the occlusion point. The options were listed in complete sentences, but contained three basic components: the hazard (e.g. “blue car”), its location (“in left lane”) and the event that occurred (“pulls into your lane”). In almost all cases the options within one clip differed by at least two of these components.

To ensure that it was not possible to guess the correct scenario from the text alone, eight volunteers, three from the UK and five from Malaysia, were given the 44 sets of scenario options (40 main clips and 4 practice clips) and asked to guess the correct answer without watching the videos. Malaysian volunteers were also asked whether they had any difficulty understanding the scenarios described, to ensure that the options were accessible to non-native English speakers. All volunteers scored at or below chance in this exercise, indicating it was not possible to guess the correct answer without watching the corresponding videos. None of the volunteers in this exercise participated in the later experiment.

To be consistent with Lim et al. (2013) which used eye tracking, the stimuli were played on a 17” 4:3 eye tracker monitor at a resolution of 1024 x 768, presented using Tobii Studio 2.3. Participants were seated 65 cm from the screen, and their eye movements were monitored using a Tobii T60 eye tracker in Malaysia and Tobii T1750 eye tracker in the UK.

2.3. Design

A 2 x 2 x 2 mixed design was used. The between-groups factors were the origin country of the participant (driver origin: Malaysia or UK) and experience level (novice or experienced). The within-groups factor was the country where the clip was filmed (country clip: Malaysia or UK).

The stimuli were separated by country into two blocks of 20 clips, i.e. one Malaysia block and one UK block. Within each block, the order of clips to be presented was randomized using a Latin square. The order of the blocks was counterbalanced across participants.

2.4. Procedure

After giving informed consent, participants completed a brief demographic questionnaire. Participants were informed that each clip contained a driving scenario leading up to a hazardous event, however the clips would end immediately before this event actually occurred and their task was to predict what the event was by selecting the correct scenario out of four possible options. They were informed that in every case, one and only one of the four scenarios had actually taken place and there was therefore a correct answer for each clip. It was also emphasized that their task was not to choose the event that they felt was the most hazardous, but the one that was most likely to have occurred.

Before each block of clips, participants attempted two practice clips, filmed in the same country as the block they were about to view. After the practice clips they were able to ask questions or seek clarification. Participants were not given any feedback as to the correct scenarios at any point during the practice clips or main experiment.

A short line of text was displayed for 1 second before each clip, indicating participants' progress through the block. After watching each clip, participants selected the scenario they thought most likely to occur by pressing the corresponding number on a numeric keypad (1, 2 3 or 4). They were then asked to rate how confident they were in their answer and how hazardous the situation was, on a 6-point scale where a higher rating indicated higher confidence or hazardousness respectively. There was no time limit imposed for participants to answer any of the three questions, and they were able to ask the researcher questions to clarify their understanding of the scenarios. After confirming their third and final answer, the progress text appeared to signal the beginning of the next clip (or end of the block, if appropriate), and the process was repeated until the end of the block. After the first block, participants were given the opportunity to take a brief break, and the process was repeated.

3. Results

Accuracy scores were analyzed using a 2 x 2 x 2 mixed ANOVA. The between-groups factors were the origin country of the participant (driver origin: Malaysia or UK) and experience level (novice or experienced). The within-groups factor was the country where the clip was filmed (country clip: Malaysia or UK). The relationship between licensing time, accuracy scores, answer confidence, perceived hazardousness, and three self-reported measures (driving ability, awareness of other road users, and general driving confidence) was then examined. Finally, chi-square goodness of fit tests were conducted for individual clips to analyze the plausibility of the incorrect, distractor options in each video.

3.1. Accuracy

Accuracy results are summarized in Figure 1. 0.5% of participant responses were deemed invalid due to incorrect keypresses and excluded. Main effects were found for all three factors: driver origin, where UK drivers outperformed Malaysian drivers ($F_{1,73} = 7.58$, $p = .007$, $\eta^2_\rho = .094$); driver experience, where experienced drivers outperformed novices ($F_{1,73} = 4.38$, $p =$

.040, η^2_ρ = .057); and country clip, where participants were more accurate on Malaysian clips ($F_{1,73}$ = 25.98, p < .001, η^2_ρ = .262). The latter two effects were subsumed by an interaction of experience and country clip ($F_{1,73}$ = 7.89, p = .006, η^2_ρ = .098). Post-hoc analyses revealed that experienced drivers outperformed novices only on UK clips (t(75) = 3.12, p = .003, d = .721) and there was no difference of experience in Malaysian clips (t(75) = .365, p = .716, d = .084). Furthermore, experienced drivers' accuracy was similar on both sets of clips (t(37) = 1.57, p = .125, d = .288), but novices were significantly better at predicting events in Malaysian clips compared to UK (t(38) = 5.40, p < .001, d = .865).

3.2. Correlational analyses

Correlations were conducted across all participants to assess whether predictive accuracy was related to a number of factors, including licensing time, two further experimental measures (the hazard and confidence ratings that participants gave for each clip), and three self-reported measures (driving ability, awareness of other road users, and general driving confidence). As there were 21 correlations in total, a Šidák-corrected α level of .0314 was used to determine significance. Results are reported in Table 1.

As expected, all three self-rated measures were strongly correlated (all *p*s < .001 and all *r*s > .500), suggesting that participants tended to rate themselves similarly on all three measures. However, only driving ability and awareness of other road users significantly correlated with participants' accuracy scores (r = .270, p = .017 and r = .253, p = .026 respectively), while self-rated driving confidence correlated with the experimental measure of participants' confidence in their answers (r = .358, p = .001), suggesting that participants appeared to exhibit similar levels of confidence in both their driving ability and answers in the clips.

Similar to Jackson et al. (2009), participants that rated clips as more hazardous were also more confident in their answers (r = .484, p < .001), suggesting that the more hazardous a person rated a clip, the more confident they were in the answer they gave. As also observed by Jackson et al. (2009), there was no relationship between participants' answer accuracy and confidence (r = .024, p = .833).

Finally, licensing time was linked with only driving ability out of the three self-rated measures (r = .300, p = .009), and also marginally correlated with accuracy (r = .220, p = .058), although this was not significant.

3.3. Distractor option plausibility

To determine whether the distractor options (i.e. the three incorrect options) were equally plausible, a chi-square goodness of fit test was performed for each clip; as there were 40 clips, a Šidák-corrected α level of .0218 was used to determine significance. Only incorrect options chosen by participants were included in this analysis; correct responses were excluded. Results are reported in Table 2.

Out of 40 clips, participants' incorrect answers were not equally distributed in 16, suggesting that for these clips, one or more of the distractor options was chosen substantially more often compared to the others. Two independent t-tests compared experienced and novice driver performance between these 16 clips and the remaining 24, and found that when the direction of performance was taken into account (i.e. whether experienced or novice drivers were more accurate), there was no difference between these two groups of clips (t(19.1) = .902, p = .378, d = .343). However, when only the magnitude of difference in performance was taken into account, there was a larger performance gap in the clips where the distractor options were not equally distributed (t(19.3) = 2.77, p = .012, d = 1.05), suggesting that when one or more of the distractor options was particularly plausible compared to the others, either experienced or novice drivers substantially outperformed the other group.

4. Discussion

4.1. Experience differentiation in UK clips

Only UK clips were found to differentiate experience, while Malaysian clips did not, and this was the case for drivers from both countries. There are several possible explanations for this difference: a ceiling effect may have occurred in the Malaysian clips, the quality of the distractor options may be superior in the UK set of clips, it may be caused by an inherent difference between the Malaysian and UK driving clips/environment, or it may be due to the nature of individual clips. We will examine these possibilities separately.

First, drivers were considerably more accurate on the Malaysian clips (68.6 %) compared to the UK clips (61.7 %). Jackson et al. (2009) observed a similar effect when using two different conditions, finding that in the more difficult condition, experienced drivers outperformed novices, but there was no group difference in the easier condition. It is therefore possible that a ceiling effect occurred, and the relative ease of the clips meant that there was no difference between experienced and novice driver performance. However, it should also be noted that 68.6% seems rather low to constitute a ceiling effect, especially given that Jackson et al. (2009) observed scores of 80% under similar conditions with a free response paradigm.

Second, the quality of the distractor options for individual hazards may play a role in experience differentiation; of the sixteen clips where incorrect answers were unevenly distributed, ten were filmed in Malaysia while only six were filmed in the UK. This suggests that when all distractor options appear equally plausible, clips may differentiate experience more successfully. Furthermore, when accuracy scores in each clip were taken into account, the performance gap between experienced and novice drivers was considerably larger in the clips where incorrect answers were skewed towards one option in particular, compared to clips where all distractor options were chosen equally often. However, this was only the case when only the magnitude and not the direction of this performance gap was taken into account. In other words, when some distractors were more plausible than others, the performance gap between the two groups widened; however, since novices outperformed experienced drivers in a small number of clips, this performance gap sometimes favored experienced drivers and sometimes favored novices. Therefore, while this is a potential consideration for future multiple choice tests, it does not appear to confer a consistent advantage for driving experience.

There may also be an inherent difference between the UK and Malaysian clips beyond simple ease of prediction, which makes the Malaysian clips less suited for differentiating experience given a multiple choice predictive paradigm. One obvious difference is that the clips used to reflect the driving environment in Malaysia were generally more hazardous and more visually cluttered (see Figures 2 and 3 for a comparison). It may be the case that the Malaysian driving environment necessitates a more even spread of attention compared to driving in the UK; in other words, a strategy that retains a high level of awareness of the various developing hazards in the environment – and therefore better readies the driver to deal with any of them – while deploying only limited attention to an immediate hazard, may be more conducive in Malaysia.

Furthermore, while it may seem counterintuitive that hazards in a visually cluttered environment would be easier to predict, especially as previous research has suggested that novices are slower to process traffic situations (Borowsky et al., 2011; Crundall & Underwood, 1998; Jackson et al., 2009), it is possible that the answer lies in the foreshadowing element of the clips (Crundall et al., 2012; Garay-Vega & Fisher, 2005). A more cluttered environment generally translates into more multiple choice options where the precursors are salient and clearly visible. Conversely, in many of the UK clips, precursors were less salient than precursors in the Malaysian clips. If all precursors are highly salient, novice and experienced drivers may be equally proficient at predicting which of these precursors becomes a hazard; however, when no precursors are particularly salient, novice drivers may fail to notice the correct precursor at all, and subsequently be less accurate at predicting hazards. Crundall et al. (2012) found that learner drivers were more likely to miss BP precursors compared to experienced drivers; as the majority of hazards fell into this category, this may reflect the results found in the present study.

However, it is unlikely that differences in precursor saliency are systematic between countries. This may also be reflected in the response data; while accuracy was higher on the Malaysian clips as a group, in both Malaysian and UK clips there were large variations in response patterns between clips. For instance, in a small subset of clips novices actually outperformed experienced drivers, while in other clips, experienced drivers were considerably more accurate. This appears to reflect previous findings that the efficacy of a hazard perception test largely depends on the clips used (Sagberg & Bjørnskau, 2006), but again, we can offer only a post-hoc explanation. Therefore, while the present study may help to provide an insight into the elements that differentiate experience, further research is required to examine these various contributing factors separately.

4.2. Cross-cultural differences in predicting hazards

Lim et al. (2013)'s previous cross-cultural study found that both novice and experienced Malaysian drivers were considerably slower to react to hazards than UK drivers, and speculated that this was largely due to a difference in hazard criterion rather than one in hazard perception ability. In a predictive paradigm such as the "What Happens Next?" test, this possibility has been eliminated, as drivers' accuracy at predicting an event should be unaffected by their opinion of its hazardousness.

However, UK drivers still outperformed Malaysian drivers on the "What Happens Next?" test, suggesting that differences in hazard perception ability may in fact exist between the driver groups, and the previous difference in reaction times was likely a combination of actual ability, thresholds for danger, and/or different ideas of what constitutes a hazard. It is also possible that UK drivers' superior performance in both studies stems from greater participation in hazard perception-type experiments and relevant training, as all UK participants would have practiced for and passed the traditional hazard perception test in order to obtain their license.

Furthermore, there was no interaction between driver origin and country clip, implying that drivers' familiarity with a location has a marginal at best effect on their ability to predict hazards. This lends further evidence to hazard perception skill being highly transferable and relatively unaffected by familiarity with an area, as UK drivers were able to predict Malaysian hazards just as accurately as they were UK hazards. It also has implications for drivers in Malaysia and potentially other developing countries, as the results suggest that a hazardous driving environment may negatively impact one's hazard perception ability, contributing at least in part to higher accident rates and possibly creating a self-perpetuating cycle. While it is uncertain exactly how significant a role hazard perception ability plays in a driver's safety compared to factors such as risk acceptance and perceptual judgment, the results of the present study suggest that driver training and road safety interventions, in both developed and developing countries, could benefit from including a hazard perception component.

4.3. "What Happens Next?": A viable hazard perception paradigm?

As discussed by Jackson et al. (2009), the "What Happens Next?" test appears to be a feasible hazard perception test paradigm, providing a more in-depth accuracy measure than reaction time. For instance, a successful response in a reaction time paradigm entails detecting a hazard in an early stage of development, while Jackson et al. (2009)'s paradigm required not only early detection but also future prediction, asking separate questions about a hazard's source ("What is the hazard?"), its location ("Where is the hazard?") and future events relating to it ("What happens next?"). Indeed, Jackson et al. (2009) found that a successful early detection did not necessarily entail an accurate later prediction, as drivers' accuracy dropped significantly with each subsequent question. Furthermore, when using a button press paradigm, one cannot be sure that participants are responding to the same hazard defined by the researchers or a different hazard altogether, although this issue has been compensated for in various ways (Lim et al., 2013; Wetton et al., 2010, 2011). The present study's results suggest that a predictive paradigm may be a more powerful differentiator of experience than response latencies, as the UK clips that failed to differentiate experience in a reaction time paradigm (Lim et al., 2013) did differentiate in the present study. Furthermore, Lim et al. (2013) found a cultural interaction where drivers identified more hazards from their own country; however, no such effect was present in this study, suggesting that a predictive paradigm may be less affected by cultural differences. If this is the case, it carries implications for drivers obtaining a license outside their home country, as they may find a reaction time test disproportionately difficult due to cultural differences rather than their level of hazard perception skill.

The predictive paradigm used in this test has now been found to differentiate experience with both a multiple choice and free response format (Jackson et al., 2009). From a practical standpoint, this has positive implications as the ease and simplicity of a multiple choice format makes this paradigm a viable option for widespread testing. Compared to the free response format, a forced choice measure has the additional advantage of controlling for participants' talkativeness, while also ensuring that they cannot respond with a null answer. From a theoretical perspective, the different formats raise several interesting possibilities; for instance, it may be the case that the Malaysian clips did not differentiate experience because the multiple choice format was relatively easy, inviting a comparison between the two formats. Furthermore, as discussed, it is possible that novice drivers' accuracy in part stems from precursors being highlighted for them; a free response paradigm with these same videos would eliminate this possibility.

Finally, a predictive task also circumvents any response bias that may exist in participants. Lim et al. (2013)'s cross-cultural study suggested that events needed to reach a higher degree of hazardousness to elicit a response from Malaysian drivers. These biases may confound the RT paradigm, as participants' responses may reflect not when they first recognize an event as a potential hazard, but rather, when it has progressed to the point that they are willing to identify it as hazardous. This is particularly relevant in developing countries with higher accident rates, where drivers are more likely to be desensitized to hazards. While it is certainly possible that an RT paradigm could be successfully calibrated, Wallis & Horswill (2007) found that manipulating test instructions had no effect on responses. This finding, combined with the results of the present study, suggests that a predictive task may be a practical alternative in countries where desensitization is likely to occur. While, as with the reaction time paradigm, the exact nature of effective clips remains to be explored in depth, these early results appear promising for hazard perception testing in developing countries.

References

Borowsky, A., Oron-Gilad, T., Meir, A., & Parmet, Y. (2011). Drivers' perception of vulnerable road users: A hazard perception approach. *Accident Analysis & Prevention*, *44*(1), 160–166. doi:10.1016/j.aap.2010.11.029

Borowsky, A., Oron-Gilad, T., & Parmet, Y. (2009). Age and skill differences in classifying hazardous traffic scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(4), 277–287. doi:10.1016/j.trf.2009.02.001

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis & Prevention*, *42*(4), 1240–1249. doi:10.1016/j.aap.2010.02.001

Chapman, P., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, *27*(8), 951–964. doi:10.1068/p270951

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention*, *45*, 600–609. doi:10.1016/j.aap.2011.09.049

Crundall, D., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, *41*(4), 448–458. doi:10.1080/001401398186937

Drummond, A. E. (2000). Paradigm lost! Paradigm gained? An Australian's perspective on the novice driver problem. Presented at the Novice Drivers Conference, Bristol, UK.

Garay-Vega, L., & Fisher, D. L. (2005). Can Novice Drivers Recognize Foreshadowing Risks as Easily as Experienced Drivers? Presented at the Driving Assessment 2005: 3rd

International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Retrieved from http://trid.trb.org/view.aspx?id=763536

Horswill, M. S., Marrington, S. A., McCullough, C. M., Wood, J., Pachana, N. A., McWilliam, J., & Raikos, M. K. (2008). The Hazard Perception Ability of Older Drivers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(4), P212–P218.

Jackson, L., Chapman, P., & Crundall, D. (2009). What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics*, *52*(2), 154–164. doi:10.1080/00140130802030714

Lim, P. C., Sheppard, E., & Crundall, D. (2013). Cross-cultural effects on drivers' hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, *21*, 194–206. doi:10.1016/j.trf.2013.09.016

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis & Prevention*, *38*(2), 407–414. doi:10.1016/j.aap.2005.10.014

Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S., & Wetton, M. A. (2011). A hazard perception test for novice drivers. *Accident Analysis & Prevention*, *43*(1), 204–208. doi:10.1016/j.aap.2010.08.010

Toroyan, T. (2013). *Global status report on road safety 2013: supporting a decade of action*. Geneva: World Health Organization. Retrieved from http://www.who.int/mediacentre/news/releases/2013/road_safety_20130314/en/index.html

Wallis, T. S. A., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis & Prevention*, *39*(6), 1177–1185. doi:10.1016/j.aap.2007.03.003

Wells, P., Tong, S., Sexton, B., Grayson, G., & Jones, E. (2008). *Cohort II: a study of learner and new drivers: volume 1* (Road Safety Research Report No. 81). London: Department for Transport.

Wetton, M. A., Hill, A., & Horswill, M. S. (2011). The development and validation of a hazard perception test for use in driver licensing. *Accident Analysis & Prevention*, *43*(5), 1759–1770. doi:10.1016/j.aap.2011.04.007

Wetton, M. A., Horswill, M. S., Hatherly, C., Wood, J. M., Pachana, N. A., & Anstey, K. J. (2010). The development and validation of two complementary measures of drivers' hazard perception ability. *Accident Analysis & Prevention*, *42*(4), 1232–1239. doi:10.1016/j.aap.2010.01.017

Figure captions

Figure 1. Hazard perception scores based on accurate predictions. Malaysia is abbreviated to MY.

Figure 2. Practice video for the Malaysian block of clips.

Figure 3. Practice video for the UK block of clips.

Table 1

Correlations for all drivers (n = 77)

	Accuracy score	Hazard rating (clips)	Confidence in answers (clips)	Driving ability	Awareness of others	Confidence in driving
Licensing time	.220	-.179	-.019	.300**	-.080	.118
Accuracy score		-.016	.024	.270*	.253*	.029
Hazard rating (clips)			.484**	.073	.038	.030
Confidence in answers (clips)				.152	.057	.358**
Driving ability					.560**	.720**
Awareness of others						.504**

*Correlation is significant at Šidák-corrected α level of .031 (two-tailed)
**Correlation is significant at the 0.01 level (two-tailed)

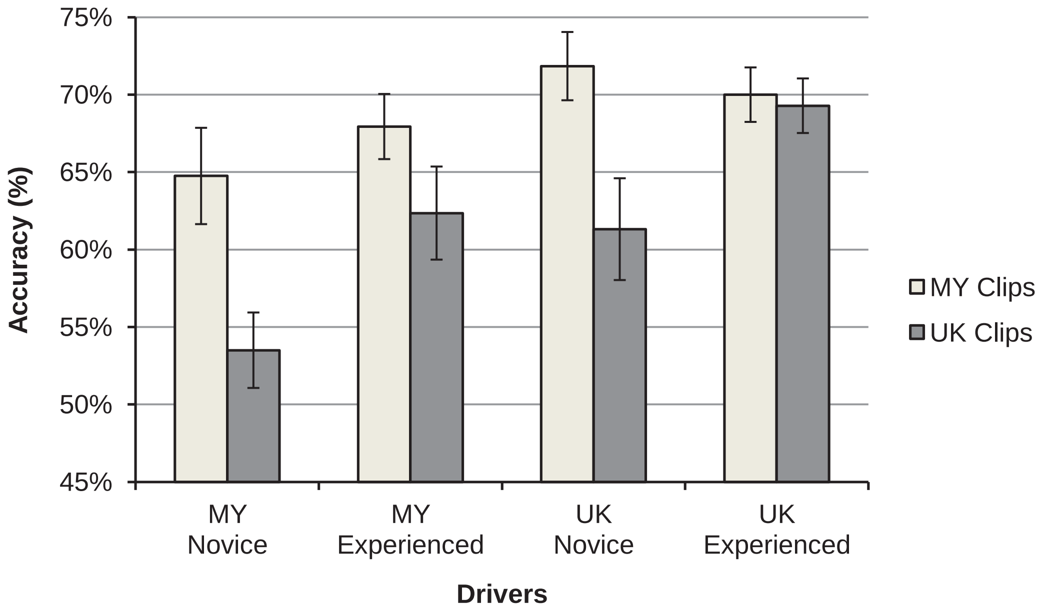
Table 2

Distribution of responses for individual clips. χ^2 analysis conducted with distractor options only

Clip category		Clip no.	Correct response	Distract or 1	Distract or 2	Distract or 3	χ^2	*p*
Malaysian clips	Matched clips	01	47	11	0	19	18.20	<.001*
		02	56	14	3	3	12.10	0.002*
		03	42	42	5	6	50.30	<.001*
		04	72	1	2	2	0.40	0.819
		05	25	8	20	23	7.41	0.025
		06	46	18	6	7	8.58	0.014*
		07	75	0	0	2	4.00	0.135
		08	41	8	6	21	11.37	0.003*
		09	66	6	2	2	3.20	0.202
		10	60	7	9	1	6.12	0.047
	Unmatched clips	11	30	30	17	7	14.78	0.001*
		12	71	2	1	3	1.00	0.607
		13	54	5	0	18	22.52	<.001*
		14	74	2	0	1	2.00	0.368
		15	57	9	10	1	7.30	0.026
		16	48	23	6	0	29.45	<.001*
		17	72	0	5	0	10.00	0.007*
		18	65	3	4	4	0.18	0.913
		19	53	3	9	12	5.25	0.072
		20	39	39	5	3	52.26	<.001*

UK clips	Matched clips	21	65	4	5	3	0.50	0.779
		22	47	1	14	14	11.66	0.003*
		23	52	3	16	5	12.25	0.002*
		24	49	7	9	11	0.89	0.641
		25	65	4	4	4	0.00	1.000
		26	66	7	0	3	7.40	0.025
		27	49	11	5	11	2.67	0.264
		28	38	18	6	15	6.00	0.050
		29	32	32	16	12	11.20	0.004*
		30	53	6	13	5	4.75	0.093
	Unmatched clips	31	61	5	10	1	7.63	0.022
		32	39	11	3	24	17.74	<.001*
		33	23	19	15	20	0.78	0.678
		34	36	23	16	1	18.95	<.001*
		35	50	13	3	10	6.08	0.048
		36	62	3	12	0	15.60	<.001*
		37	37	17	10	13	1.85	0.397
		38	60	10	3	4	5.06	0.080
		39	34	7	15	21	6.88	0.032
		40	47	9	9	11	0.28	0.871

* Significant at Šidák-corrected α level of .021


75%
70%
65%
60%
55%
50%
45%
Accuracy (%)
MY Novice
MY Experienced
UK Novice
UK Experienced
Drivers
MY Clips
UK Clips