Predicting criminal incidents on the basis of non-verbal behaviour: the role of experience

David Crundall & Lauren Eyre-Jackson

Division of Psychology

School of Social Sciences

Nottingham Trent University, UK

Address for correspondence:

Prof. David Crundall

Division of Psychology

School of Social Sciences

Nottingham Trent University,

Burton Street

Nottingham NG1 4BU

e-mail: david.crundall@ntu.ac.uk

Keywords: situation awareness, criminal behaviour, prediction

ABSTRACT

Do experienced police officers have a superior ability to detect impending criminal acts? In order to test this hypothesis ten Closed Circuit Television (CCTV) clips were collected from real criminal events that occurred in and around Nottingham city centre in the UK. Ten control clips were filmed specifically or chosen from existing footage to match the criminal clips, but did not contain any criminal activity. All clips ended abruptly, immediately prior to a real criminal act unfolding, or a non-criminal act in the control clips, and either the screen turned black, masking the video scene, or remained frozen on the final frame of the edited clip. Thirty police officers and 30 control participants watched the clips. At the end of each clip, participants were asked to predict what would happen next. Signal detection analysis indicated marginal evidence that police show greater accuracy in predicting clips that cut to black screen compared to the general public. A stronger effect was noted in the analysis of the criterion, with police officers much more likely to predict a crime regardless of whether there was one. These findings provide promising evidence of experiential differences between police officers and the general public when identifying criminal and antisocial behaviour in CCTV footage, though the greater criterion bias effect suggests that experience may over-sensitise individuals to non-verbal cues.

INTRODUCTION

Experience in a particular environment can interact rapidly with early visual processing of a scene. Within 150 ms participants can extract enough information about the *gist* of a scene to allow them to classify it into a broad category without actually identifying any objects within that scene (Torralba, Oliva, Castelhano and Henderson, 2006). This early information may be based on low-level spatial frequencies that is quickly extracted and compared to stored representations of environments (or averages, or prototypes of environments that have been built up over multiple exposures), which can then restrict goal-directed visual search to pertinent areas of the visual scene, and prime compatible object recognition (Cheung and Bar, 2012). However all representations of scenes will be based on experience within those scenes. If one has never viewed a kitchen, one might have no representations to guide one to find the toaster. Conversely, if one happens to be a highly experienced kitchen-fitter, then interpretation of the visual scene is likely to be much more nuanced. If the task was to locate an integrated (concealed) refrigerator, one might predict that the expert kitchen-fitter would use subtler cues to identify its location sooner than an average observer. Similarly domain expertise can improve the ability to both spot and then identify certain objects: bird watchers are more likely to spot and correctly identify birds, just as motorcyclists are more likely to notice motorcycles on the road, and then be able to identify the make (e.g. Crundall, Clarke, Ward and Bartle, 2008; Tanaka and Taylor, 1991; Johnson and Mervis, 1998). Johnson and Mervis (1998) argued that domain experts use different features to categorise objects: "Novices' categories are based on overall similarities among objects, including shape, size, and color…, whereas experts' categories are based on more abstract criteria, such as behavioural similarity…", p385. Once specific objects have been identified or categorised by comparison with stored representations, or through the use of heuristics (such as the 3+1 toe pattern of song birds), associative links prime the probability of other objects being nearby (Bar, 2003; 2004). Thus spotting the fork on one side of the plate, suggests that the knife should be on the other side. It is rare however that experience of a context is limited to static snapshots of reality. We experience sequences of events: just as the bottle of champagne might suggest the nearby location of champagne glasses, so the act of the waiter opening the champagne bottle leads to the prediction that the wine will be subsequently poured. Through these associative links, we thus predict events. Again, domain experience will lead to more accurate predictions of future events, sometimes without conscious understanding of how these predictions are made. For instance Klein, Calderwood and Clinton-Cirocco (1986) used the Critical Decision Method to identify the cues that firefighters use to identify the source location of a fire. The expert firefighters identified several cues, including the colour of the flame and the movement of smoke, which they reported had never previously been discussed or trained.

When placed within a security context, one might thus expect that trained and highly experienced police, CCTV operators, prison officers, etc. should demonstrate superior skills in predicting criminal behaviour on the basis of subtle cues. Following Johnson and Mervis (1998) the experts may be more likely to classify potential criminals by subtle behavioural cues behaviour, rather than on the basis of "shape, size and color". For instance, the avoidance of gaze contact might indicate deception, or slight non-verbal behaviour might allow one to predict an imminent violent outburst. Despite this logical extension of experiential effects in scene perception and behavioural prediction to the security services, the reported evidence in favour of such a hypothesis is mixed at best (e.g. Troscianko et al., 2004; Garrido, Masip and Herrero, 2004; Johnson, 2007).

Research on police officers’ attempts to identify deception has noted that they rely on cues that have repeatedly been found to be non-predictive of deception in experimental studies (Akehurst et al., 1996; Stromwall and Granhag, 2003; Vrij and Semin, 1996). One such cue, explicitly taught in the Reid Method of Interviewing and Interrogations (Blair and Kooi, 2004), is that deceivers will often avoid gaze contact, yet there is experimental evidence to suggest that liars will maintain gaze contact more so than truth-tellers (e.g. Walcyzk et al., 2012).

While there has been limited investigation of police officers abilities to predict violent or criminal behaviour from immediate behavioural cues, there have been a number of studies conducted with CCTV operators. Again, logical extensions of theoretical arguments would suggest that highly trained and experienced CCTV operators would have a greater chance of predicting whether a social interaction is likely to develop into a criminal or anti-social incident. In an assessment of training practices for CCTV operators, Darker et al. (2007) reported that there was very little formal training in the detection of suspicious behaviour, with many operators or control-room managers emphasising the role of on-the-job experience built up over time. One manager also commented that ex-store security guards often become good CCTV operators because they have already have experience in identifying suspicious behaviour. Thus it appears that anecdotal evidence from CCTV staff and managers also support the hypothesis that experience is crucial to predicting criminal activity on the basis of non-verbal cues. However, as with studies of police officers’ skills in detecting deception, the experimental evidence does not support this view.

Perhaps the most oft-quoted study that attempted to find an experiential benefit of CCTV operators in predicting criminal acts was undertaken by Troscianko et al. (2004). They presented one hundred 15-second clips of CCTV footage to a group of CCTV operators and a control group of students, with instructions to rate the clips for their likelihood of leading to a criminal or antisocial event. Eighteen of these clips contained footage that preceded real criminal events (e.g. assault), though they stopped prior to the criminal act and paused on the final frame for a further 5 seconds. A further 18 clips were selected as controlled matches for the criminal clips, containing similar settings and scenarios. The overall correct classification rates of the incident and matched control clips were 69.5% and 72% for the CCTV operators and students, respectively. While a signal detection task (SDT) analysis suggested that participants were able to successfully predict whether or not clips would lead to a criminal incident, no evidence was found for greater sensitivity in the CCTV operators (though the authors did report the CCTV operators to have a more liberal criterion in identifying clips as leading to a criminal event, regardless of whether or not they did so).

More recently, Wijn, van der Berg and Lousberg (2013) compared CCTV operators of high and low experience (though only the average level of experience across the two groups is reported). Using a similar freeze-frame methodology to that of Troscianko et al (2004) they also failed to find an experiential difference. Similarly, Blecho, Darker and Gale (2005) could not find any experiential advantage when they tested the ability of 8 CCTV operators and 8 lay people to identify whether targets in mock CCTV footage were carrying concealed weapons (as opposed to concealed non-offensive objects).

The most promising attempt to identify an experiential benefit was reported by Grant and Williams (2011). They presented 24 clips to 12 CCTV operators and 12 control participants. Following Troscianko et al.'s methodology, half of the clips ended immediately prior to a criminal act while the other half ended without leading towards an incident. Following the presentation of each clip, participants provided ratings regarding potential criminality. They found that CCTV operators correctly classified 55.5% of the incident and control clips, while the control participants only correctly classified 46.5% of clips. However when subjected to a multivariate analysis of variance (including many other dependent variables), the experiential factor was not found to be significant, though it remains a possibility that, had an SDT approach been taken, a comparison of sensitivity (d') across groups may have revealed an experiential effect.

One difference between the study of Troscianko et al., (2004; and also Wijn et al., 2013) and that of Grant and Williams (2011) is that, while not explicitly discussed, the method section of Grant and Williams' paper suggests that the clips simply finished before ratings were required, instead of pausing on the final frame as in Troscianko et al.'s study. We believe this otherwise innocuous difference may be key to identifying a potential experiential benefit in correctly classifying criminal and control clips prior to an incident taking place, and may also provide greater understanding of the experiential mechanism for improving performance. If one presupposes that experience may guide the eyes of the observer to prioritise the most appropriate areas of the scene (e.g. the antagonist) *at the most appropriate time* (e.g. just before a punch is thrown), then using a freeze-frame methodology at the end of each clip would potentially remove any experiential benefit. While experts might look in the right place at the right time in order to predict subsequent behaviour, freezing the final frame gives the control participants ample time to view the static scene and spot the crucial clues that they would have otherwise missed if the event had unfolded in real time.

This identical issue was reported in studies of driving expertise. Vogel et al. (2003) found that highly experienced drivers were no better than novices at predicting the development of traffic scenarios from filmed clips that froze on the final frame. Jackson, Chapman and Crundall (2009) argued that the final frozen frame provided more visual information than novices drivers would normally have. While the experienced drivers might process enough information for accurate prediction prior to the clip freezing, novices could still extract the same level of information at a later stage from the frozen image.

To test this, Jackson et al. (2009) compared novice and experienced drivers on their ability to predict 'what happens next?' in driving clips that either froze on the final frame, or were suddenly occluded. The results demonstrated that experienced drivers performed better than the novices when the clip ending was occluded. Novices' performance on the freeze-frame condition was however comparable to the experienced drivers performance in the occluded condition. Jackson et al. (2009) interpreted this as evidence that the benefit of experience in predicting driving scenarios is, at least in part, derived from timely interrogation of the visual scene. When novice drivers were no longer under any processing speed pressure (in the freeze frame condition) this experiential benefit disappeared, as the novices could view the visual cues for prediction at their leisure.

This sudden-occlusion methodology is a form of Situation Awareness Global Assessment Technique (SAGAT; Endsley, 2000) that is often used to probe participants' Situation Awareness in a variety of operational settings (e.g. air traffic control). Grant and Williams may have come closest to this methodology, though they did not analysis the sensitivity of their participants to identify whether their apparent differences were significant. Key to this application of the SAGAT method is the need to "stop the clip at a pre-determined time unknown to the participant and ask questions to measure knowledge about the current and future situation" (Jackson et al., 2009, p157). This criterion is not met in either the Troscianko et al. study or that of Grant and Williams, where both studies used clips that were consistently 15 seconds in length. This consistent length may have provided a temporal warning for all observers that may have artificially increased novice attentiveness in the final seconds of each clip, further masking any experiential differences.

The current study mirrors that of Jackson et al. (2009) using CCTV clips that are immediately occluded following presentation or that remain frozen on the final frame. We predict that experienced observers of criminal and antisocial events (police officers in the current study) would be more sensitive to whether a clip preceded such an event, when compared to a control group of participants, providing the clips are immediately occluded rather than frozen on the final frame.

METHOD

Participants. Thirty police officers and thirty members of the general public were recruited to take part in this study. Police officers were recruited from a police station in the south of England and were unfamiliar with the locations depicted in the stimuli. This group comprised of 19 males and 11 females, with a mean age of 36 years (SD= 8.98, range 21-56). The mean number of years' service in the police was 10.38 (SD= 7.04, range 0.5-28 years). Members of the general public were recruited from local businesses in the same area that police participants were recruited, and were also unfamiliar with the locations featured in the video clips. This group comprised of eleven males and nineteen females. The mean age of this group was 35.76 (SD=11.97, range 20-63 years).

Design. This study used a 2x2x2 mixed design. The between-subjects variables were *group* (*police* vs. *public*), and *video display,* with each clip ending either with the last frame visible for 3 seconds, or with an immediate occlusion by a black screen (*freeze frame* vs. *occluded*). The within-subjects variable was *criminal content* with half of the clips ending immediately prior to a criminal act, while the other half did not (*crime* vs. *control*). The dependent measures were the percentage of clips correctly identified according to their criminal content, reaction times for making this decision (recorded from the end of the 3 second occlusion/free frame), accuracy in predicting the type of criminal activity that might take place (from a choice of 4 options), and a self-confidence rating for predictions (made on a 7-point scale, with 1 reflecting extremely low confidence and 7 reflecting extremely high confidence).

Materials & Apparatus. Ten short clips of real CCTV footage were obtained with the assistance of Nottingham City Council. Criminal acts included physical assault, theft and criminal damage. These clips were edited to end immediately prior to the criminal act (e.g. the footage might show an animated discussion between two males, and may even include body posture clues immediately prior to a physical attack, but would end just before the first punch was thrown). Ten control clips were edited from similar footage from the same cameras used in the control clips or from similar locations. In several instances, control clips were taken from the actual original footage that led to the criminal incident, though control footage was extracted prior to the instigators of any criminality arriving on the scene. This method of creating control clips ensured that the setting and even some of the people were similar to the crime clips, though there was never any duplication of footage across clips. Other clips were selected from footage that mirrored the criminal footage in terms of setting (time of day, location, camera position) and pedestrians in the scene (e.g. one control clip contained a group of men walking through a pedestrianized area in the day time wearing similar clothes to the men in the corresponding crime clip). Figure 1 contains the final frame of a crime clip and its corresponding control clip.

All twenty clips had time and date stamps blurred out. Nottingham City Council protocol required individual faces of a certain size (a 12% screen threshold) to be blurred. For those crime clips that required facial blurring, we ensured that the corresponding control clip also contained facial blurring. All clips involved panning and zooming of the cameras but did not contain any audio. Some crime videos were black and white (using a night filter) while others were in colour. Matching control clips copied the format of their paired crime clip wherever possible. All clips had a resolution of 720 x 576 and subtended approximately 9 x 7 degrees of visual angle at a viewing distance of 40 cm.

To create the two levels of *video display*, clips were edited into two separate formats: the final frame was immediately replaced by a black screen for 3 seconds (the *occluded* condition), or the final frame remained on screen for 3 seconds following the end of the clip (the *freeze frame* condition).

Following a participant's decision as to whether the clip was leading to a criminal activity or not, four alternative crimes were then presented, and the participant was asked to choose one (regardless of whether or not she thought the clip was going to lead to a crime, or whether the clip actually led to a crime). The four alternative options were derived from a focus group who viewed the crime clips and provided alternative criminal endings. The options that were identified for each crime clip were also paired with the corresponding control clips. Options included assault, arson, indecent exposure, theft etc.

The experiment was run on a laptop using Psycho-Py2.

[insert figure 1 here]

Procedure. Prior to testing, each participant was randomly assigned to one of the two video display conditions (freeze frame or occluded). Members of the general public took part in the experiment in a quiet location of their choice (e.g. office or home) and police participants were tested at their station. Participants were initially asked to sign a consent form and fill in a demographic questionnaire including age, gender, and job title. They were then instructed that they would see 20 clips filmed from CCTV cameras located in Nottingham and that it was their task to determine whether each clip would lead to a potentially criminal incident. At the end of each clip, following a three second pause filled by either a black screen or a freeze frame of the final clip frame, they were prompted to answer this question by pressing either 'Y' or 'N' on the keyboard. Participants where then asked – "assuming that an incident was about to occur" - which of four criminal incident options was the most likely to occur. Participants responded by pressing 'A', 'B', 'C', or 'D' on the keyboard corresponding to their chosen criminal incident. Finally participants were asked how confident they felt about the option they had just chose on a scale of 1 to 7 where 1 = not confident and 7 = very confident. This Likert scale could be completed by pressing an appropriate number key on the keyboard or by using the trackpad to select an option with the cursor arrow. Before starting the block of trials, a practice trial was given and participants had the opportunity to ask questions. After the experiment participants were thanked for taking part and debriefed, but did not receive any payment for taking part.

RESULTS

Accuracy for discriminating criminal content

All sixty participants completed the study and provided full data sets for the analyses. An initial 2 x 2 x 2 mixed Analysis of Variance (ANOVA) was conducted to compare accuracy at determining whether a clip would lead to a crime or not across the factors of group (police vs. public), display type (occluded vs. freeze frame) and criminal content of the clips (crime vs. control). A significant main effect of criminal content was noted ($F(1,56)=6.80$, $MSe=270.71$, $p<0.05$) with participants correctly identifying crime clips more so than control clips (73% vs. 65%). This is highly suggestive of a bias across participants to report that a crime is about to occur, regardless of whether the clip precedes an actual criminal incident. This main effect was however subsumed by a significant interaction between criminal content and group ($F(1,56)=6.23$, $MSe=270.71$, $p<0.05$). The interaction (see Figure 2) clearly demonstrates that the main effect for correctly identifying crime clips at the expense of the no crime clips is solely driven by the police group. Two post-hoc t-tests (with Bonferroni corrections) however revealed while the police were better than the public at identifying the crime clips ($t(58) = 3.0$, $p < 0.005$), the corresponding dip in accuracy for no-crime clips was not sufficient to differentiate them from the public ($t(58) = 1.2$). Thus while there is a suggestion for a potential criterion bias in the police's willingness to label an event as potentially criminal, there is also evidence of increased accuracy compared to the general public. In order to untangle police sensitivity for detecting criminal events from a potential bias towards reporting all events as criminal, individual measures of sensitivity (d') and criterion (c) were calculated.

[insert figure 2 here]

Sensitivity and criterion bias for discriminating criminal content

Sensitivity (d') and criterion (c) were calculated for each participant according to Stanislaw and Todorov (1999). Two participants were more than 2.5 standard deviations away from the sample mean (one with poor sensitivity, and another with extreme bias towards reporting a crime). These two participants (one from the occluded condition and one from the freeze frame condition; both members of the general public) were removed from all further analyses.

The calculated measures of d' for correctly classifying a clip as leading to a crime were subjected to a 2 x 2 between-subjects ANOVA across the factors of participant group (police vs. public) and video display (freeze frame vs. occluded). There was no main effect for group of video display ($Fs(1,54)<1$). The results however indicated marginal evidence for a significant interaction between group and display condition ($F(1,54)=3.23$, $MSe=1.45$. $p=0.078$). Figure 3a shows that the police and general public do not show a difference in accuracy for the freeze frame condition, but there is an ostensible increase in sensitivity for police in the occluded condition as predicted.

[insert figure 3 here]

The criterion measures for each participant were also included in a 2 x 2 between-groups ANOVA across group and video display. There was a statistically significant main effect for group ($F(1,54)=11.24$, $MSe=0.41$, $p<0.05$) with police officers having a greater likelihood of reporting a crime compared to the general public (Figure 3b). Neither the main effect for display condition ($F(1,54)=0.23$, $MSe=0.41$, $p=0.63$) nor the interaction effect ($F(2,54)=0.02$, $MSe=0.41$, $p=0.88$) reached statistical significance.

Response times to discriminating criminal content

A 2 x 2 x 2 mixed ANOVA on the response times for participants (with the two outliers removed) revealed a statistically significant main effect for clip type ($F(1,54)=12.81$, $MSe=1.24$ $p<0.05$) with participants being quicker to respond correctly to crime clips (mean=2.33s) than control clips (mean=3.07s). There was also a statistically significant interaction for criminal content and group ($F(1,54)=3.94$, $MSe=1.24$, $p=0.05$; see Figure 4) suggesting police to have the fastest response times to a correctly identified crime clip, yet the slowest response times to a correctly identified clip where no crime takes place. Despite the interaction, t-test comparisons with Bonferroni corrections failed to confirm group differences at either level of criminal content. Nonetheless the pattern is highly similar to that noted with the analysis of percentage accuracy.

Identifying the potential crime

Following the simple crime/no crime response required immediately after the end of the clip, participants were then asked to identify what that crime might have been (even if they had just responded that the previous clip did not lead to a crime). Though all clips were followed by 4 potential crimes to choose from, only the crime clips had a correct response. Accuracy for picking the correct crime for the 10 crime clips was analysed via a 2 x 2 between groups ANOVA across participant group and video display. A main effect of participant group was found ($F(1,54) = 4.2$, $MSe = 125.4$, $p<0.05$) with police outperforming the general public (55.7% vs 49.6%). The video display (occluded vs. freeze frame) produced neither a main effect ($F(1,54) < 1$) nor an interaction with group ($F(1,54) = 2.8$).

Following the choice of a crime, participants were asked how confident they were in their choice. Confidence scores (from a 1 to 7 scale, with 7 being highly confident) were subjected to a 2 x 2 x 2 ANOVA, resulting in a main effect of criminal content ($F(1,54) = 90.9$, $MSe = 0.25$, $p < 0.001$) and an interaction between criminal content and participant group ($F(1,54) = 4.9$, $MSe = 0.25$, $p < 0.05$). As can be seen in figure 5, the confidence for both groups following crime clips is identical, but the interaction is driven by the greater drop in confidence for police compared to the public when faced with no crime clips.

DISCUSSION

The results have identified that police participants are better than control participants at identifying the type of crime that is about to be committed, and there is marginal evidence that they are more sensitive to the imminent possibility of a criminal or anti-social act. The analysis of d’ supports the hypothesis that the occluded condition is most likely to evoke the experiential difference. When control participants are given additional time to process the final image in the freeze frame condition, their sensitivity for criminal content becomes almost identical to that of the police officers. This supports the suggestion that the benefit of experience in predicting criminal or anti-social outcomes is partly derived from prioritising elements in the visual scene both spatially and temporally. The control participants are less likely than the police officers to be looking in the most informative location at the point of occlusion. Thus in the freeze frame condition, control participant performance is likely to approach that of the experienced observers, because their ability to interrogate the visual scene is not time limited.

Interestingly, while the occlusion condition in the current study seems to have had an effect on responses regarding whether or not there is going to be a crime (at least in the suggested interaction of the sensitivity measure), there was no ostensible effect of occlusion in deciding what the crime might be. This suggests that information which allows participants to decide on a subsequent crime type is available earlier that the information required to judge whether a crime is actually about to happen. This is effect will no doubt be heavily influenced by the nature of the 3 distracter options that were provided along with the correct answer. For instance, if the clip contains only one person it reduces the possibility that the resultant crime would involve a physical assault. The distraction options were however chosen by a focus group who were tasked to identify credible distracter items through discussion, though it remains a possibility that all options were not as equally distracting.

The other key finding from the current study was the more liberal criterion employed by the police officers in deciding whether a crime or anti-social act was imminent. This may reflect the greater likelihood of police officers to witness such incidents in the course of their job, which may bias them towards a criminal classification when all other elements are held constant. Alternatively, the result may simply reflect the costs associated with false positives and false alarms that are learned with on-the-job experience (i.e. police may prefer the lesser costs associated with many false alarms than the potentially high costs associated with a single false positive). Troscianko et al. (2004) also noted a criterion difference between their groups: though their experts showed little bias, their control participants were more conservative (experiment 1).

Though the current results can be used to criticise the methodology of previous studies which have used freeze frame presentations, there is the possibility that the marginal evidence for expert sensitivity might be more to do with the choice of our expert group. The current study assessed the abilities of police officers instead of CCTV operators as used by Troscianko et al. (2004; and also Wijn et al., 2013; Grant and Williams, 2011). Could police officers have better predictive skills than CCTV operators, thus providing an alternate explanation for the marginal effect on sensitivity? This would accord with Darker et al.'s (2007) recorded observation that ex-store security guards make good CCTV operators because of their real world experience. Certainly any real incidents that police observe are more likely to be witnessed in person rather than over a CCTV network. The increased salience and threat of these incidents may have a greater impact on the observations skills of on-the-scene police rather than the CCTV operators who are removed from the situation. However, testing police officers' abilities to predict criminal behaviour using stimuli presented from a CCTV operator's perspective would appear to add an extra layer of extrapolation and complexity. Thus we argue the greater mapping between CCTV operators experiences and the CCTV perspective used in the current clips might actually increase the experiential gap should operators be compared to control participants using this occlusion methodology. Unfortunately we did not have access to a naive group of CCTV operators to test this additional hypothesis (most operators in Nottingham had already seen many of the incident clips used in this study), though it provides a distinct avenue for future research.

In conclusion, the results suggest that experiential benefits for police officers predicting crimes from CCTV clips might indeed be identifiable, contrary to extant evidence, providing the clips are occluded prior to the incident. The promising use of occlusion suggests that the benefit of expertise lies in not just prioritising the most appropriate areas of the scene for visual attention, but also attending to them at the most appropriate time.

ACKNOWLEDGMENTS

We would like to thank Kevin Bond and all the staff at Woodlands Surveillance Control Room for their assistance in obtaining the stimuli. We would also like to thank all the police officers and participants who took part in this study.

REFERENCES

Akehurst, L., Köhnken, G., Vrij, A., and Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology, 10,* 461-471.

Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience, 5*, 617-629.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15,* 600-609.

Blair, J. and Kooi, B. (2004). The gap between training and research in the detection of deception. *International Journal of Police Science and Management, 6, 2,* 77-83.

Blechko, A., Darker, I. T., and Gale, A. G. (2008). Skills in detecting gun carrying from CCTV. Proceedings of the 42nd IEEE International Carnahan Conference on Security Technology, 265-271.

Crundall, D., Clarke, D., Ward, P., and Bartle, C. (2008). *Car drivers' skills and attitudes to motorcycle safety.* London: Department for Transport.

Cheung, O. S, and Bar, M., (2012). Visual prediction and perceptual expertise. *International Journal of Psychophysiology, 83,* 156-163.

Darker , I., Gale, A., Ward, L., and Blechko, A. (2007). Can CCTV reliably detect gun crime? Proceedings of the 41st IEEE International Carnahan Conference on Security Technology, 264-271.

Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: A critical review. In: M.R. Endsley & D.J. Garland (Ends), Situation Awareness Analysis and Measurement. Mahwah, NJ: Lawrence Erlbaum Associates.

Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology, 39,* 254-275.

Grant, D., and Williams, D. (2011). The improtnance of perceiving social contexts when predicting crime and antisocial behaviour in CCTV images. *Legal and Criminological Psychology, 16,* 307-322.

Jackson, L., Chapman, P., and 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.

Johnson, K. E., and Mervis, C. B. (1998). Impact of intuitive theories on feature recruitment throughout the continuum of expertise. *Memory and Cognition, 26, 2,* 382-401.

Johnson, R. R. (2007). Race and police reliance on suspicious non-verbal cues. *Policing: An International Journal of Police Strategies and Management, 30, 2,* 277-290.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting, 1*, 576–580.

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments and Computers, 31, 1,* 137-149.

Stromwall, L. and Granhag, P. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology Crime & Law, 9, 1,* 19-36.

Tanaka, J.W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

Torralba,A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review, 113, 4,* 766.

Troscianko, T., Holmes, A., Stillman, J., Mirmehdi, M., Wright, D., and Wilson, A. (2004). What happens next? The predictability of natural behaviour viewed through CCYV cameras. *Perception, 33,* 87-101.

Vogel, K., et al., 2003. Traffic sense – which factors influence the skill to predict the development of traffic scenes? *Accident Analysis and Prevention, 35, 5,* 749–762.

Vrij, A. and Semin, G. (1996). Lie experts' beliefs about non-verbal indicators of deception. *Journal of Non-verbal Behavior, 20, 1*, 65-80.

Wijn, R., van den Berg, H., and Lousberg, M. (2013). On operator effectiveness: the role of expertise and familiarity of environment on the detection of deviant behaviour. *Personal Ubiquitous Computing, 17,* 35-42.

Walczyk J. J., Griffith D. A., Yates R., Visconte S. R., Simoneaux B., Harris L. L. (2012). Lie detection by inducing cognitive load: eye movements and other cues to the false answers of false answers of "witnesses" to crimes. *Criminal Justice and Behavior, 39,* 887–909.

FIGURE TITLES

Figure 1. The final frame from a crime clip (top panel) and the corresponding control clip (bottom panel). The crime in question is a physical assault: in the top panel a man in black approachings the group from the top of the scene and pulls his arm back ready to throw a punch. The control clip is filmed from the same camera on a different date, but at a similar time of night, with similar individuals passing through the scene.

Figure 2: Percentage accuracy for deciding whether ended immediately prior to a criminal act across participant group and actual criminal content within the clips (with standard error bars added)

Figure 3: Sensitivity (top panel) and criterion (bottom panel) for police and the general public when identifying clips as either leading to a crime or not (with standard error bars added)

Figure 4: Response times to correctly identified trials (with standard error bars added).

Figure 5: Confidence scores for selecting the type of crime across participant groups and the criminal content of the clip (with standard error bars added).

Figure 1.

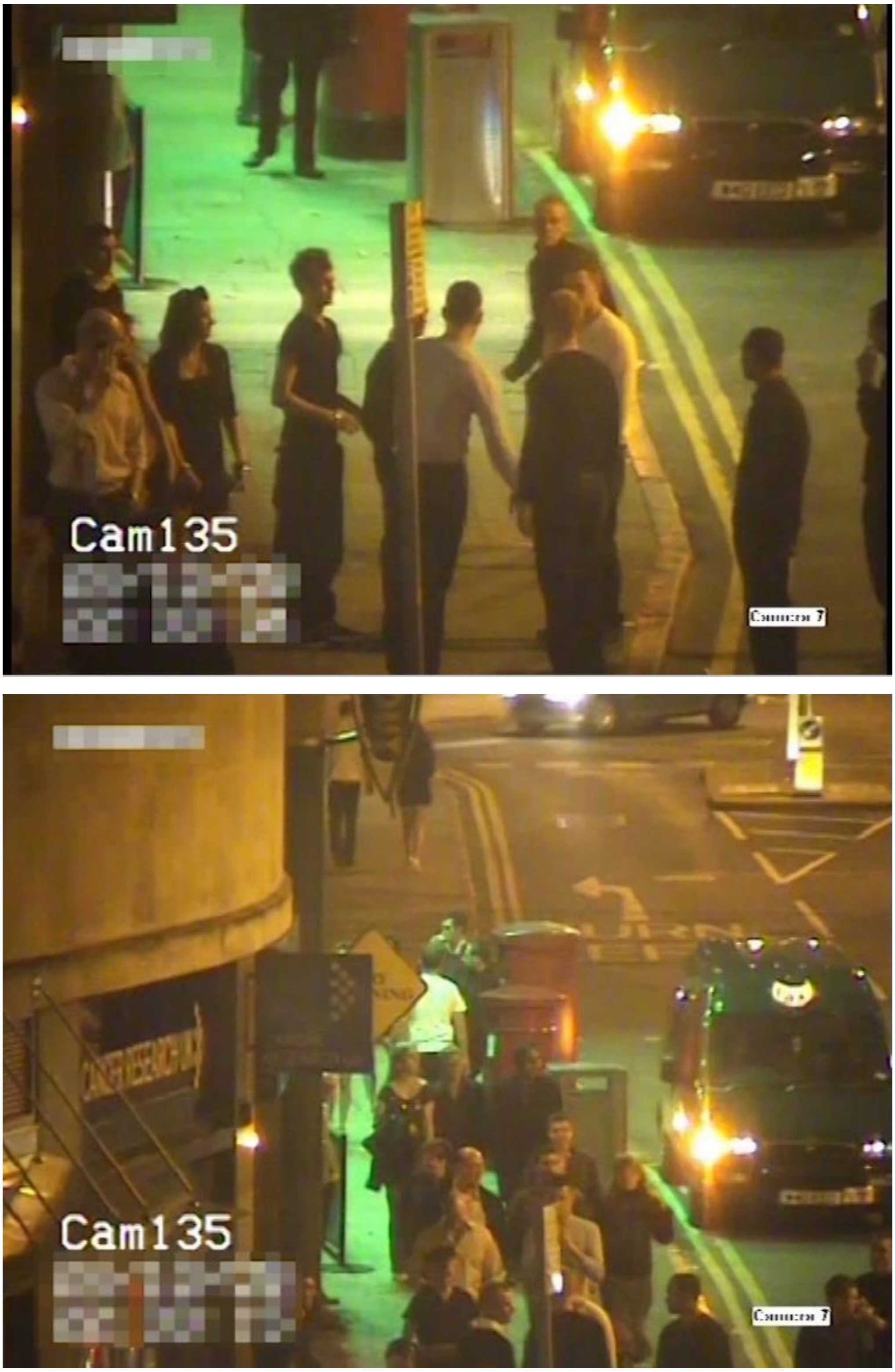


Figure 2.

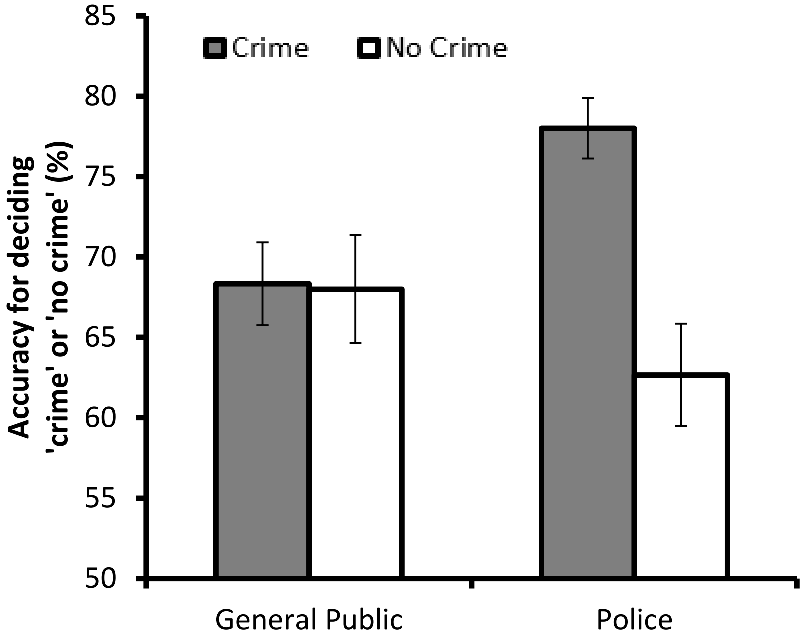
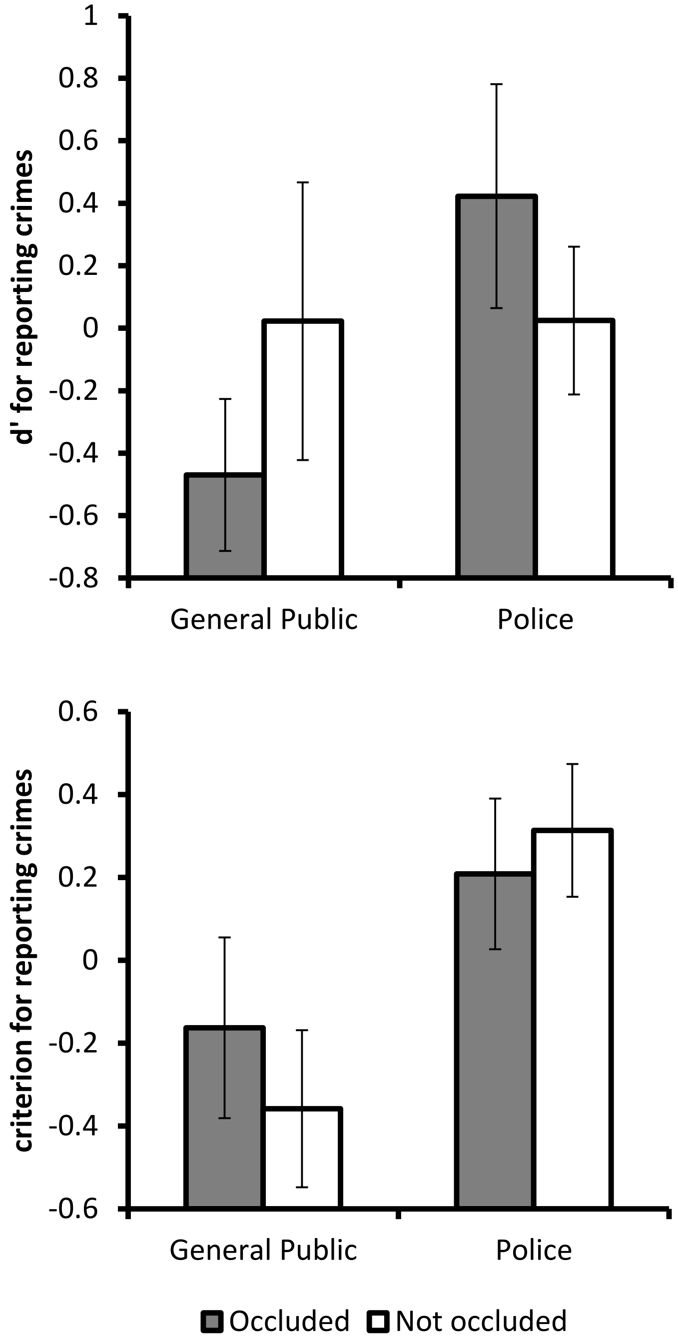

85
80
75
70
65
60
55
50
Accuracy for deciding 'crime' or 'no crime' (%)
Crime
No Crime
General Public
Police

Figure 3.



1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
d' for reporting crimes
General Public
Police
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
criterion for reporting crimes
General Public
Police
Occluded
Not occluded

Figure 4.

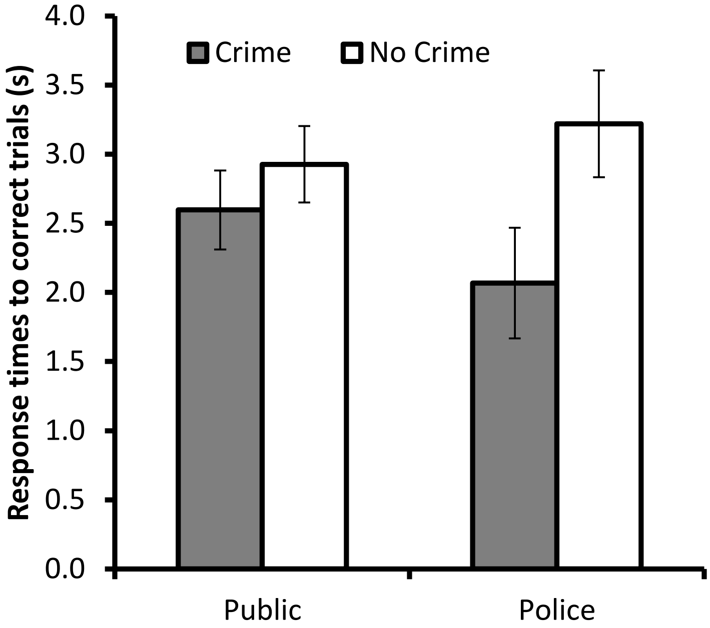
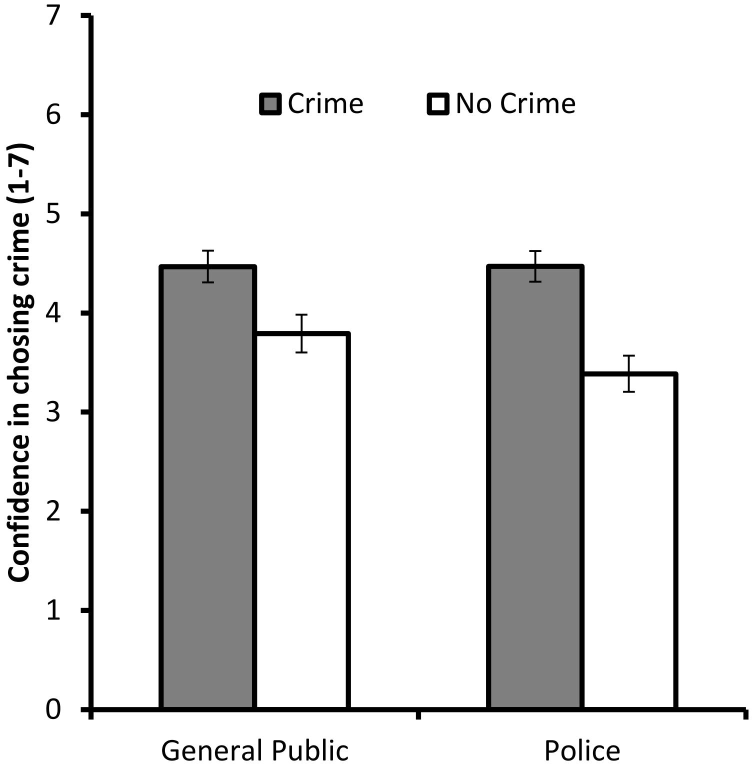

4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Response times to correct trials (s)
Crime
No Crime
Public
Police

Figure 5.


Crime
No Crime
Confidence in chosing crime (1-7)
0
1
2
3
4
5
6
7
General Public
Police