Perception, 2004, volume 33, pages 87 – 101

DOI:10.1068/p3402

What happens next? The predictability of natural behaviour viewed through CCTV cameras†

Tom Troscianko, Alison Holmes, Jennifer Stillman¶, Majid Mirmehdi, Daniel Wright§, Anna Wilson§

Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK; ¶ School of Psychology, Massey University, Albany, North Shore MSC, Auckland, New Zealand; § Department of Psychology, University of Sussex, Falmer, Brighton BN1 9QH, UK; e-mail: tom.troscianko@bris.ac.uk

Received 1 July 2002, in revised form 15 September 2003

Abstract. Can potentially antisocial or criminal behaviour be predicted? Our study aimed to ascertain (a) whether observers can successfully predict the onset of such behaviour when viewing real recordings from CCTV; (b) where, in the sequence of events, it is possible to make this prediction; and (c) whether there may be a difference between naïve and professional observers. We used 100 sample scenes from UK urban locations. Of these, 18 led to criminal behaviour (fights or vandalism). A further 18 scenes were matched as closely as possible to the crime examples, but did not lead to any crime, and 64 were neutral scenes chosen from a wide variety of noncriminal situations. A signal-detection paradigm was used in conjunction with a 6-point rating scale. Data from fifty naïve and fifty professional observers suggest that (a) observers can distinguish crime sequences from neutral sequences and from matches; (b) there are key types of behaviour (particularly gestures and body position) that allow predictions to be made; (c) the performance of naïve observers is comparable to that of experts. However, because the experts were predominantly male, the absence of an effect of experience may have been due to gender differences, which were investigated in a subsidiary experiment. The results of experiment 2 leave open the possibility that females perform better than males at such tasks.

1 Introduction

We explored the question whether people can predict if an antisocial, or violent, act is about to be carried out by an individual or group of individuals, in real urban settings, as observed by CCTV (closed-circuit television) cameras. The literature reveals that observers can identify important information about others from purely visual cues. This ability has been the focus of considerable research. Asch (1946) said that people have the capacity to "understand something about the character" of others (page 258). For example, Ambady and Rosenthal (1993) found that students, teachers, and strangers, could reliably predict the effectiveness of teachers from very brief silent video clips. Other video studies have demonstrated the importance of nonverbal behaviours to third-party observers in judgments of managerial effectiveness (Burnett and Motowidlo 1998), extraversion (Lippa and Dietz 2000; Riggio et al 1990), intimacy (Floyd 1999), and rapport (Grahe and Bernieri 1999). People can also recognise intentional acts (for a review, see Baldwin and Baird 2001). With minimal information, for instance when, as with dynamic point-light displays, the human visual system has access solely to movement information, individuals can identify, for example, the human figure (Johansson 1973), its gender (Kozlowski and Cutting 1977), emotional state (Dittrich et al 1996; Walk and Homan 1984), types of actions (Dittrich 1993), vulnerability to attack (Gunns et al 2002), and even the gait of friends (Cutting and Kozlowski 1977; Dittrich et al 1994). Recent psychophysical studies have also shown the benefits of motion-based studies in face recognition. Utilising state-of-the-art motion capture and animation techniques Hill

† Parts of this research were previously presented in a paper given at the 24th European Conference on Visual Perception, Kuşadasi, Turkey.

and Johnston (2001) found that head motion, as opposed to motion of facial features, provided useful information for recognising faces, and established that facial movements were beneficial in identifying gender. This motion effect has also been obtained with video images of famous faces degraded by blurring and pixelation (Lander et al 2001). However, to-date, although a large number of motion-based psychophysical studies have been conducted, we are not aware of any other ecologically valid psychophysical study (real people and real scenarios) carried out to investigate human observers' ability to predict criminal acts on the basis of the behaviours of human figures prior to those acts.

This topic has an interesting application, because there is increasing use of CCTV systems in urban areas throughout the world. One question that may be raised about their use is whether a CCTV operator is capable of knowing whether something 'bad' will happen soon in a given location, so as to be able to alert the security services to a possible risk.

In CCTV control rooms, the outputs from the surveillance cameras are typically observed by control-room staff who then alert the security services if they detect a crime or antisocial act being committed; they also monitor situations they believe will result in such acts. These judgments of intentionality are based on visual information only. Mostly, the cues lie in the sequence of actions that are visible, although there may be occasions when an individual or individuals are suspected of unlawful behaviour because they are known offenders. From our observations of several control rooms in the UK, it became apparent that an operator often monitors a large number of (up to 50) TV screens simultaneously. This perceptual load is higher than an individual can cope with (Tickner and Poulton 1973), and therefore there is considerable interest in exploring the feasibility of an automated alerting system for possible violent/antisocial acts. Tickner and Poulton also suggested that females may perform better than males in identifying incidents. This is a topic we address in experiment 2.

We carried out psychophysical experiments on human observers to ascertain whether they were able to make accurate judgments about what would happen later in scenes recorded by urban CCTV cameras. If the evidence suggests that prediction is possible, then the feasibility of producing a computational detector of the relevant behaviour is increased. The aims of this feasibility study were threefold. First, we wanted to investigate whether observers could predict criminal or antisocial behaviour from CCTV surveillance footage, and, second, whether naïve and experienced observers would differ in their judgments. Finally, and most importantly, we wished to examine whether there were identifiable behaviour patterns that might be predictive of unlawful acts.

2 Experiment 1

2.1 *Method*

2.1.1 *Participants.* Fifty professional CCTV control-room operators, currently employed in monitoring urban CCTV surveillance cameras, and unfamiliar with the locations depicted in the video clips, were recruited as our 'expert' group. This group comprised operators aged between 20 and 62 years (mean = 38.7 years), with surveillance experience ranging between 0.25 and 17 years (mean = 5.7 years). There were forty-one males and nine females. The mean age of the males was 39.32 years (SD = 11.68 years). The mean age of the females was 36.11 years (SD = 7.36 years). First-year and second-year psychology undergraduates at the University of Bristol made up the 'novice' group. They were aged between 18 and 21 years (mean = 19.1 years) with forty-two females and eight males. The mean age of the females was 19.63 years (SD = 0.83 years). The mean age of the males was 19.63 years (SD = 0.74 years). Participants were tested individually in sessions lasting approximately 1 h 45 min. Experts were paid £7.50; novices earned a course credit.

In order to protect the privacy of the members of the public who appeared in the tapes, confidentiality agreements were signed by all participants. To minimise the chances of any individual on the tape being recognised, care was taken to test in locations outside of those used for the experimental tapes.

2.1.2 *Materials.* The first stage of the study involved gaining access to a large amount of real-life CCTV footage from Police and Local Authorities and constructing the experimental tapes. Once assembled, the experimental tapes contained 100 video clips, each clip lasting 15 s. Of these, 18 were classified as 'incidents': scenes leading up to unlawful acts, such as assaults, car crimes, or vandalism. In all of these, the clip ended (by freezing) before any violent/antisocial event occurred. There were 18 'matches': scenes not leading to unlawful acts but matched, as closely as possible, to the incident clips with respect to principal characters' behaviours, age, dress style, type of location, and time of day. Examples of matches are as follows: For an assault match—it is night time on a street scattered with a number of young males. A casually dressed man, standing with a small group, notices another man in the middle of the street. He runs towards him and puts his arms across the man's torso, then ... pause (freeze). What actually happened after the pause was that the casually dressed man hugged the second man and walked back to the group with his arm around his shoulders. For a car-crime match—a man is seen standing alone for the full 15 s, next to a car in a car park then ... pause (freeze). The man was, in fact, waiting for a companion, who was off-screen buying a parking ticket. The appendix gives brief descriptions of the characters, locations, and events, up to and after the pause, for both incident and match clips. Because the majority of activity in public spaces is fairly mundane, whether it is being monitored by CCTV operators or viewed by the general public, the incident and match clips were embedded in a set which also included 64 'neutral' scenes. These depicted a variety of everyday scenes, for example people chatting, waiting at bus stops, or walking on the street. This design made possible the use of rigorous psychophysical methodology based on signal detection theory (SDT) to evaluate the *sensitivity* of an observer who is asked to judge whether a scene will lead to a crime or not. It also allows the *response bias* of the observer to be computed separately. (Harvey 2003 provides a useful summary of SDT and its use in applied settings.) If the judgments had been scored simply in terms of the percentage of correct detections of incidents, the observer's sensitivity and response bias would be confounded.

The videotapes containing the stimuli were carefully constructed to minimise the possibility of experimental artifacts. Each scene played for 15 s then paused for 5 s on the last frame. To avoid order effects, 9 of the 18 incidents and 9 of the 18 matches were shown at random (Lindley and Scott 1984) in the first 50 clips, and then mirrored in the following 50. For example, video clips 1 and 100 were neutral clips, and clips 2 and 99 were both incident clips leading to assaults. Two experimental tapes were assembled with the presentation order reversed. These two tapes were shown equiprobably to the observers, so that half of the observers saw tape 1 and half saw tape 2. Participants viewed the tapes on a 21-inch (54 cm) television screen from a distance of 1 m. The bottom 10 cm of the screen were masked in order to remove cues contained within the date and time markers. After an initial pilot study, which confirmed that the instructions were sufficient to enable participants to complete the tasks, the study proceeded to testing. The *Perception* website http://www.perceptionweb.com/misc/p3402) contains two example video stimuli. One is an 'incident' example, and leads to the woman in the white top punching another person. The 'match' does not lead to any aggressive behaviour. Faces have been blanked out for the purposes of publication. The 'incident' example results are given in figure 3.

2.1.3 *Procedure.* The participants' task was to judge whether something 'bad' had occurred in the moments following what was shown on each video clip. They were told that the definition of something bad was something unlawful or antisocial, and that, in each case, their response should relate only to the characters and/or situations on the screen. A 6-point rating scale was provided, with the following verbal definitions: 1, certain not bad; 2, fairly certain not bad; 3, unsure, probably not bad; 4, unsure, probably bad; 5, fairly certain bad; 6, certain bad. To familiarise participants with the procedure a short rehearsal, with an 8-scene tape, was conducted. A further summary of the instructions was then provided prior to the experimental tape. Responses were recorded in an experimental booklet.

2.2 *Results and discussion*
2.2.1 *Incidents and matches.* Prior to a detection-theory analysis, we undertook a simple comparison of the responses of the experts and novices to the 1800 incident and 1800 match video clips (each, 18 scenes × one-hundred observers). Figure 1 graphically represents the outcome of the judgment task for each group. For this comparison, ratings of 4, 5, and 6 to incident clips, and ratings of 1, 2, and 3 to match clips, were scored as correct. When broken down by group, the proportion correct for incidents scenes (hits) was as follows: experts 0.82, and novices 0.79; and the proportion correct for match scenes (correct rejections) was: experts 0.57, and novices 0.65.

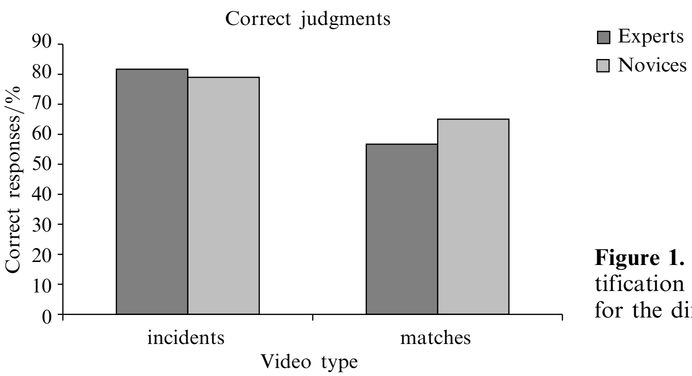

Figure 1. Probability of a correct identification of crime and match events for the different groups of observers.

2.2.2 *Sensitivity and bias.* In keeping with detection theory, the perceived magnitude of the evidence in favour of an incident, consequent upon the presentation of each video clip, is assumed to fall on a continuous distribution. The mean magnitude of these internal representations is assumed to be higher for those arising from the incident clips than from those arising from non-incident or match clips. Some overlap between the two distributions is assumed. The magnitude of the evidence required for a response in each of the 6 rating categories is assumed to increase from a rating of 1 (certain not bad) to a rating of 6 (certain bad). The bias index specifies the location of the boundary between the evidence required for a response in a particular category, and that required for a response in an adjacent category. With 6 response categories there will be 5 category boundaries, or criteria.

To obtain sensitivity and bias estimates, the observers' ratings were pooled and used to construct receiver operating characteristics (ROCs). The usual SDT model applied to data such as these is the normal–normal model, often assuming equal variances. However, in this study a logistic model was adopted because it provided a better fit to the data. The adoption of a logistic, over a Gaussian, model did not alter the pattern of intergroup relationships evident in the results. For either model, the χ^2 goodness-of-fit statistics establish that the distribution associated with the incident clips had a greater variance than distributions associated with either the remaining clips, or with match clips.

Therefore an unequal-variance model was selected. From the ROCs, estimates of various parameters were obtained by means of a maximum likelihood algorithm (RSCORE 5.3.2; Harvey 2001). These included the sensitivity parameter d_a'; the area, A_z, under the ROC; and the category boundaries, c_a (refer equation 3.13, page 74, Macmillan and Creelman 1991). d_a' is appropriate to unequal-variance models, and is the distance between the means of the underlying signal-plus-noise (incident) and noise (either non-incident or match) distributions, in terms of the root-mean-square average of their SDs (Macmillan and Creelman 1991). We explored whether observers were sensitive to the difference between incidents and all other scenes, some of which shared elements in common with incident clips (ie match clips), and also to the particular differences between incidents and matches. Table 1 presents the results of the analyses and shows the parameter d_a' and the area, A_z, under the best-fitting ROC, together with the χ^2 goodness-of-fit statistic and the associated probability of obtaining a χ^2 this large, or larger, if the model is true. Green and Swets (1966, pages 45ff) showed that the area under the ROC is equal to the proportion correct that would be attained by an unbiased observer, in a two-alternative forced-choice task, with the same stimuli. It is therefore a measure of sensitivity that makes intergroup differences comprehensible in terms of actual performance expectations. The obtained d_a' and A_z values indicate that observers could distinguish incidents from the everyday scenes which dominate most ongoing video surveillance, and could also distinguish between incidents and matches. The sensitivity measures are slightly higher for the novices than the experts in both analyses. Nevertheless, the standard errors of the two groups are barely separated for incident versus all non-incident judgments, and overlap for match versus incident judgments. The performance of the two groups should thus be considered comparable. The slightly poorer fits to the combined data may, nevertheless, suggest some differences between the groups, that must be accommodated by a compromise when the data are combined.

Table 1. Experiment 1. Logistic unequal-variance model maximum-likelihood estimates (±SD) of d_a' and A_z, the area under the ROC. The χ_3^2 goodness-of-fit statistics, with associated p-values, the bias measures, c_a, and the number of trials contributing to the estimates are also given.

	Number of trials	d_a'	A_z	χ^2	p	c_a
Incidents versus all non-incidents						
all participants	10000	1.60 ± 0.034	0.89 ± 0.005	8.63	0.03	0.026
experts	5000	1.52 ± 0.047	0.88 ± 0.007	3.27	0.35	−0.060
novices	5000	1.70 ± 0.049	0.90 ± 0.007	6.19	0.10	0.125
Incidents versus matches						
all participants	3600	1.06 ± 0.037	0.80 ± 0.008	5.54	0.14	−0.26
experts	1800	1.02 ± 0.052	0.78 ± 0.011	2.89	0.42	−0.36
novices	1800	1.10 ± 0.052	0.80 ± 0.011	3.67	0.30	−0.19

The analysis program (RSCORE) reports the category boundaries, c_a, separating the responses associated with each of the ratings. For the unequal variance case, c_a specifies the boundary in SD units relative to the zero-bias location. The criterion separating ratings of 3 and 4 also separates all ‘not bad’ responses (ratings 1 – 3), from all ‘bad’ responses (ratings 4 – 6). This central boundary was chosen to characterise the groups' biases. A value of zero indicates no response bias, values less than zero indicate a tendency, when uncertain, towards reporting an incident, and values above zero indicate a tendency, when uncertain, towards reporting a non-incident. Table 1 shows that, when judging incident clips in the context of all other clips, the experts demonstrated little response bias, while the novices, despite their equivalent sensitivity, were somewhat more conservative, tending to require more evidence before reporting an incident. With respect to judgments of incidents versus matches, all groups tended to be more inclined to decide in favour of reporting an incident than not, with the experts' response criterion continuing to be a little more liberal than the novices. Given the low a priori probability of an incident amongst all other scenes ($p = 0.18$), it is not surprising that the observers were more conservative in their judgments than they were when judging incidents versus matches, for which the probability of an incident was higher at $p = 0.5$.

2.2.3 *Experts.* To further examine the experts' performance when the more challenging incident-versus-match judgments were called for, they were segregated into three groups on the basis of experience: either low, medium, or high. The low-experience group had seventeen members (five female, twelve male) mean age 29.9 years (SD = 8.1 years) with between 0.25 and 2.6 years experience (mean = 1.5; SD = 0.8 years); the medium group had eighteen members (two female, sixteen male) mean age 29.9 years (SD = 8.1 years) with between 3 and 7.5 years experience (mean = 4.9 years; SD = 1.6 years); and the most experienced group had fifteen members (two female, thirteen male) mean age 48.7 years (SD = 6.8 years) with 8 to 17 years experience (mean = 11.4 years; SD = 3.0 years). The proportion correct for incidents (hits) was: 0.81, 0.81, and 0.83 for the groups with low, medium, and high experience, respectively. The proportion correct for match scenes (correct rejections) was: low-experience group 0.54, medium-experience group 0.57, and high-experience group 0.60 (see figure 2). The same method as above was used to construct ROCs for each group. The resulting bias and sensitivity parameters, with the associated goodness-of-fit statistics, for judgments of matches versus incidents are reported in table 2. The obtained values indicate that all of the groups could distinguish between incidents and matches. Despite a slight trend for increasing sensitivity with increasing experience, the overlap of the SDs confirms that the performances of the groups were comparable. From table 2 it can be seen that a tendency to err on the side of caution, evident in the bias index reported for experts in table 1, does not disguise any marked differences between the various experience levels, although

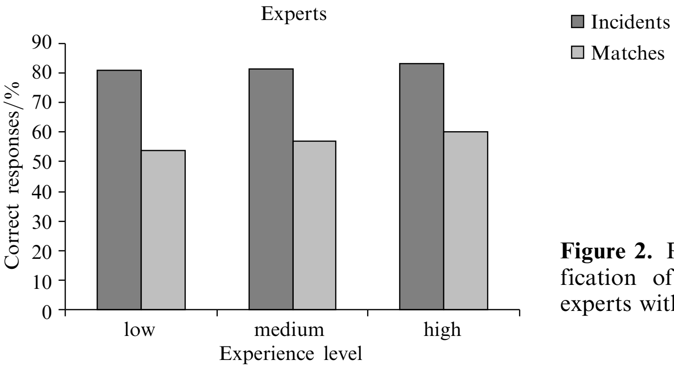

Figure 2. Probability of a correct identification of crime and match events for experts with different levels of experience.

Table 2. Experiment 1. Logistic unequal-variance model maximum-likelihood estimates (±SD) of d'_a and A_z, the area under the ROC—for participants grouped by level of experience. The χ^2_3 goodness-of-fit statistics, with associated p-values, and the bias measures, c_a, are also given.

Experience	d'_a	A_z	χ^2	p	c_a
Low	0.96 ± 0.088	$0.77 + 0.020$	1.30	0.73	−0.36
Medium	1.05 ± 0.087	$0.79 + 0.019$	3.31	0.35	−0.32
High	1.06 ± 0.095	$0.80 + 0.020$	5.12	0.16	−0.30

there may be a slight tendency for the degree of bias to be inversely related to the degree of experience. In general, all groups are somewhat more inclined to report an incident than not.

2.2.4 *Predictive behaviours.* The level of sensitivity demonstrated in the analyses suggests that there are predictive behaviours in the video clips that are recognised by experts and novices alike. We therefore attempted to identify these. To this end, incidents and matches judged "yes" (ie that an unlawful act would follow the pause) were replayed to participants. If, when they viewed these video clips, there was a point at which they decided something bad would happen, they were instructed to signal this by calling out "now". The points identified by the observers were plotted for each video clip and, provided they were not at the end of the clip, the behaviours at points with the highest predictive power were examined. In this context, 'highest predictive power' was defined as locations on the video clip which gave the highest ranked probabilities of a "now" response *and* a correct identification of the type of crime about to be committed. The following data from those video clips that produced the most powerful effects suggest that participants were recognising particular aspects of behaviour as being predictive of the onset of antisocial or unlawful acts. For example, ninety-seven participants correctly judged the incident represented in figure 3 (appendix, incident 6). Of these, eighty-five correctly predicted what actually happened after the tape had paused. The most frequently signalled point in the tape was frame number 37. At frame 37 the principal character was moving with a distinct gait whilst pointing, at frame 34 she was standing 'hands on hips'.

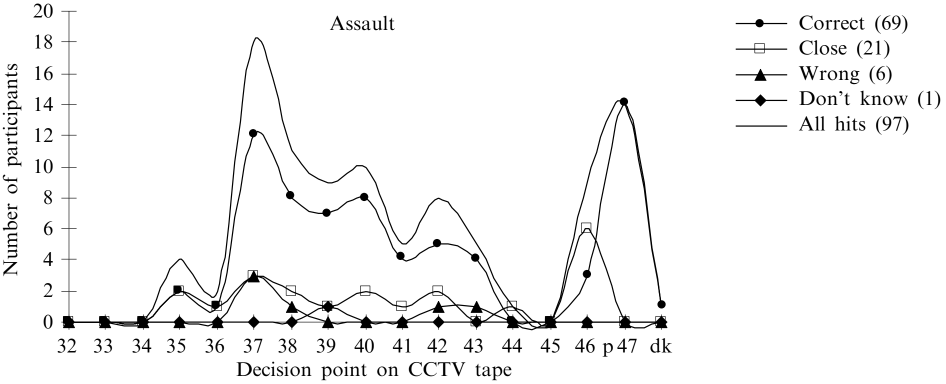

Figure 3. Example of an incident where ninety-seven participants correctly judged that an incident was about to occur. Of these, eighty-five correctly predicted what actually happened after the tape had paused. (Numbers on x axis relate to numbering from original CCTV tapes, 'p' indicates the pause and 'dk' indicates 'don't know'.) This example may be viewed in the 'incident' video on the journal website. The timing numbers are visible in the video.

Distinctive gaits and hand gestures were also features in data peaks for other incidents as was looking away from walking direction and looking around repeatedly. Figures 4a–4d (appendix, incidents 1, 7, 11, and 14, respectively) show further examples of such data peaks, and table 3 describes the behaviours at the most frequently signalled frames.

2.2.5 *Accuracy.* Finally, to investigate if participants could accurately predict what happened next, they were shown the "yes" incident and match clips again. They were told to watch the entire clip and, at the end, report what they thought happened next, ie immediately after the pause. They were informed that responses should be as specific as possible and include only characters and/or situations on the screen. They were not asked to justify the predictions and told it was unnecessary, if they did.

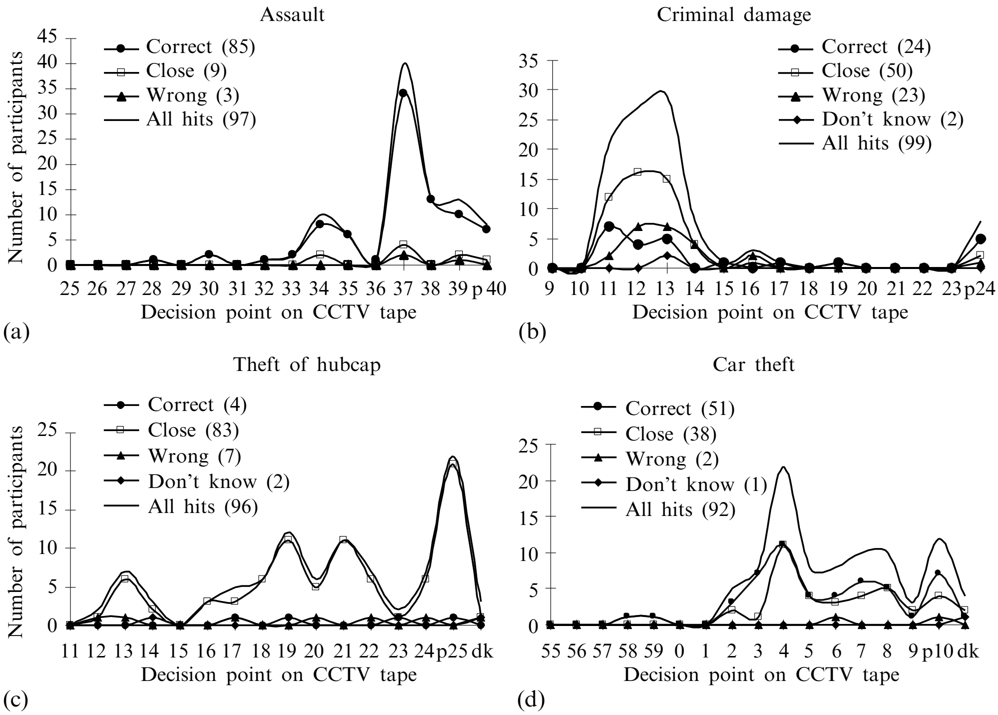

Figure 4. Examples, showing incident type, of the point at which observers decided that a crime would be committed in 15 s video sequences. (Numbers on x axis relate to numbering from original CCTV tapes, 'p' indicates the pause and 'dk' indicates 'don't know'.)

Table 3. Behaviours suggesting most predictive power and frame numbers from incident clips.

Figure	Incident type	Frame(s)	What's happening
3	Assault/fight	37	aggressive stance; gait; pointing
		34	hand gestures, hand on hip
4	Assault	37	leaves group, walking, gesturing
		40, 42	walking, gesturing
5	Criminal damage	11	pulling shirt off; aggressive gait
		12, 13	aggressive gait
6	Theft of hubcap	19, 21, 13	looking around repeatedly
7	Car theft	4, 7, 3	glancing back whilst walking forward

Table 4. Prediction accuracy, distributions for all participants, experts, and novices for video clips correctly judged as incidents.

	Correct (%)	Close (%)	Incorrect (%)	Don't know (%)	Total (%)
All participants	490 (34%)	444 (31%)	440 (30%)	72 (5%)	1446 (100%)
Experts	244 (34%)	217 (30.5%)	217 (30.5%)	33 (5%)	711 (100%)
Novices	246 (34%)	227 (31%)	223 (30%)	39 (5%)	735 (100%)

Predictions were recorded in an experimental booklet by the researcher. Accuracy rates were then coded to reflect responses as being correct, close, or incorrect. A description was classified as 'correct', if a participant was specific about both the principal character(s) and the subsequent event. A 'close' classification was made when the prediction mentioned the perpetrator(s) but was less specific about what actually happened next. For example, in an assault-type incident clip: "female in white top hits other female", or "fight, white top hits first" were classified as correct whereas "... white top going to fight with someone" was classified as close. Another professional coded the data again, and after discussion and sight of the incident clips on disputed cases unanimous agreement was reached. Table 4 shows that 65% of predictions were either correct or close, 30% were wrong, and in 5% of the cases, participants were unable to make a prediction; they just felt something 'bad' was going to happen. The table also shows that there was no real difference in performance between experts and novices.

3 Experiment 2: Investigation into possible gender differences

The results of the main study suggested that there was no difference between the expert and naïve participants. However, it should be noted that the experts were predominantly male, and the naïve subjects predominantly female. This bias reflects the natural populations in the two environments (control rooms and university, respectively) so we felt that the issue of gender effects should be addressed directly. To achieve this, we ran a scaled-down version of the main experiment on participants divided equally across males and females, with the two groups being balanced for the other important variables (age and experience).

3.1 *Method*

Fifty participants, twenty-five male and twenty-five female, were recruited from the student population at the University of Sussex. The mean age of the males was 22.04 years (SD = 3.05 years); the mean age of the females was 21.92 years (SD = 3.08 years). A subset of the original stimuli was used, and only yes/no confidence data were recorded. There were 8 'incident' clips, 8 matches to these, and 29 neutral clips. Thus, the prior probabilities of a "yes" response were kept similar to those in the main study.

3.2 *Results*

Table 5 shows the results of this experiment. It can be seen that the sensitivity indices are higher for the females than for the males; however, the standard errors of the estimates barely fail to overlap. In addition, the model is a poor fit to the data from the male observers. These results do, however, leave open the possibility that the absence of a difference between the experts and the naïve participants in experiment 1 may have been due to an imbalance in the genders of these two groups.

Table 5. Experiment 2. Logistic unequal-variance model maximum-likelihood estimates ($\pm$SD) of d_a' and A_z, the area under the ROC; for participants grouped by gender. The χ_3^2 goodness-of-fit statistics, with associated p-values, and the bias estimates, c_a, are also given.

	d_a'	A_z	χ^2	p	c_a
All participants	1.07 ± 0.080	0.80 ± 0.017	16.03	0.001	−0.49
Males	0.97 ± 0.109	0.77 ± 0.024	19.36	0.000	−0.43
Females	1.22 ± 0.116	0.83 ± 0.023	5.12	0.163	−0.55

4 General discussion

We carried out psychophysical experiments with human observers, to see whether they were able to make accurate judgments about what would happen later in scenes recorded by real CCTV cameras. The outcome of a signal-detection analysis established that the estimates of sensitivity were very similar for novices and experts. Overall, the model was an excellent fit to the data, and the comparability of these measures confirms that, for these participants, there were no differences in the ability of novices and experts to make the required judgments. This outcome was somewhat unexpected. A possible reason for it might be that the cues are 'automatic', low-level, ones that are not strongly modified by experience. In addition, it may be that gender differences contributed to the finding because the experts, by contrast with the naïve observers, were mostly male. On the basis of experiment 2, and in keeping with the suggestions of Tickner and Poulton (1973), there remains a possibility that females are more sensitive to the requisite cues than males. It may be that other differences between the student novices and the experienced security personnel in particular age, and degree of education, might also mask a small effect of experience. However, it seems relatively safe to assume that the effect of experience was not large.

The findings from this study suggest that prediction of future events is indeed possible, although, as would be expected, is imperfect. The area under the ROC curves, representing the detection of incident scenes amongst both ordinary and incident-like scenes, indicates that were simple categorical judgments, incident or not incident, called for, the participants would have correctly detected about 89% of incidents. The probability of an incident was low at $p = 0.18$, but clearly the incidents were distinguishable from everyday interactions. Were the participants asked to judge amongst equally probable incidents and incident-like but innocuous scenes, they would have correctly detected about 80% of incidents. We have identified key behaviours that seem to be associated with correct predictive responses. These behaviours often involve a certain kind of motion either of the whole body, or limbs, or the head. The study allows us to be optimistic that, in the future, it may be possible to design automatic units to detect such aspects of biological motion, although many more examples of CCTV clips will need to be studied before this becomes possible.

Further work is needed to establish more precisely the nature of the cues that allow correct prediction to be made. Our study has identified maximally predictive portions of the video sequence, and we have given a (verbal) description of what happens in these portions. Many of these would seem to indicate that *dynamic* cues, such as a particular gait or arm gesture, appear to predict later antisocial behaviour. The point-light technique of Johansson (1973) is useful in that it isolates purely dynamic cues—so our kind of study will need to be carried out with point-light stimuli. In general, further work will need to establish the relative strength of such dynamic versus static form-based cues (such as body posture). It would also be desirable to be able to control the shapes and movements of the forms on the screen; this would mean constructing animations based on real video stimuli, and then modifying these. The main value of the present study is to have developed a technique that tells us where to look for the most relevant cues, whatever their exact nature turns out to be. This technique will be useful for the other types of stimulus material mentioned above.

Acknowledgments. This research was supported by an EPSRC grant numbered GR/M97312. The authors gratefully acknowledge Sedgemoor District Council's Community Services Department, Thames Valley Police, and the London Borough of Newham's Emergency and Security Services, for the support and assistance that made this study possible.

References

Ambady N, Rosenthal R, 1993 "Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness" *Journal of Personality and Social Psychology* **64** 431 – 441

Asch S E, 1946 "Forming impressions of personality" *Journal of Abnormal and Social Psychology* **41** 258 – 290

Baldwin D A, Baird J A, 2001 "Discerning intentions in dynamic human action" *Trends in Cognitive Sciences* **5** 171 – 178

Burnett J R, Motowidlo S J, 1998 "Relations between different sources of information in the structured selection interview" *Personnel Psychology* **51** 963 – 983

Cutting J E, Kozlowski L T, 1977 "Recognition of friends by their walk" *Bulletin of the Psychonomic Society* **9** 353 – 356

Dittrich W H, 1993 "Action categories and the perception of biological motion" *Perception* **22** 15 – 22

Dittrich W H, Churchill A, Weidenbacher H, 1994 "Recognition of friends by different movements under point-light conditions" UH-Psychology Reports 1.7, University of Hertfordshire, Hatfield, Hertfordshire, UK

Dittrich W H, Troscianko T, Lea S E G, Morgan D, 1996 "Perception of emotion from dynamic point-light displays represented in dance" *Perception* **25** 727 – 738

Floyd K, 1999 "All touches are not created equal: Effects of form and duration on observers' interpretations of an embrace" *Journal of Nonverbal Behavior* **23** 283 – 299

Grahe J E, Bernieri F J, 1999 "The importance of nonverbal cues in judging rapport" *Journal of Nonverbal Behavior* **23** 253 – 269

Green D M, Swets J A, 1966 *Signal Detection Theory and Psychophysics* (New York: John Wiley & Sons)

Guardian, 2001 "Eye Spy", 1 August

Gunns R E, Johnston L, Hudson S M, 2002 "Victim selection and kinematics: A point-light investigation of vulnerability to attack" *Journal of Nonverbal Behavior* **26** 129 – 158

Harvey L O Jr, 2001 "RSCORE 5.3.2" Department of Psychology, University of Colorado, Boulder, Colorado

Harvey L O Jr, 2003 "Living with uncertainty in an uncertain world: Signal detection theory in the real world", in *Visual Perception. The Influence of H W Leibowitz* Eds J Andre, D A Owens, L O Harvey Jr (Washington, DC: American Psychological Association) pp 23 – 41

Hill H, Johnston A, 2001 "Categorizing sex and identity from the biological motion of faces" *Current Biology* **11** 880 – 885

Johansson G, 1973 "Visual perception of biological motion and a model for its analysis" *Perception & Psychophysics* **14** 201 – 211

Kozlowski L T, Cutting J E, 1977 "Recognising the sex of a walker from a dynamic point-light display" *Perception & Psychophysics* **12** 575 – 580

Lander K, Bruce V, Hill H, 2001 "Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces" *Applied Cognitive Psychology* **15** 101 – 116

Lindley D V, Scott W F, 1984 *New Cambridge Elementary Statistical Tables* (Cambridge: Cambridge University Press)

Lippa R A, Dietz J K, 2000 "The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments" *Journal of Nonverbal Behavior* **24** 25 – 43

Macmillan N A, Creelman C D, 1991 *Detection Theory: A User's Guide* (Cambridge: Cambridge University Press)

Riggio R E, Lippa R, Salinas C, 1990 "The display of personality in expressive movement" *Journal of Research in Personality* **24** 16 – 31

Tickner A H, Poulton E C, 1973 "Monitoring up to 16 synthetic television pictures showing a great deal of movement" *Ergonomics* **16** 381 – 401

Walk R D, Homan C P, 1984 "Emotion and dance in dynamic light displays" *Bulletin of the Psychonomic Society* **22** 437 – 440

Appendix

Time/place	Incidents		
	event description	scene at pause	after pause
1 Night/street	Middle-aged man gesticulating towards mixed group of young people. Two females leave group move towards him, one pointing; he backs away.	Female, arm raised, pointing towards man.	She assaults him.
2 Night/empty street	Two young men, one runs off pavement towards traffic bollard.	Young man with bollard in foreground.	Kicks/smashes bollard.
3 Night/empty street	Man, 30ish, walking along street.	Reaches rubbish sacks.	Kicks rubbish across street.
4 Night/street	Five young men walking; gather close.	One man leans into group.	Assaults young man on inside.
5 Day/car park	Full 15 s, lone man stands beside car.	Standing at passenger door, starts to move.	Tries to enter vehicle.
6 Night/outside nightclub	Mixed group, teens/20s; female looks around, breaks away from group and moves, pointing, towards another female.	Reaches her, arm raised.	She assaults her.
7 Night/street	40ish man walks away from mixed group. Strips off t-shirt whilst walking.	Has reached pub with clenched fist.	Punches pub window.
8 Night/car park	Man, 20s, crouching between two cars. Passes something into passenger door of one. Crouching, takes something from passenger.	Crouching, leans towards adjacent car.	Steals hub-cap from adjacent car.
9 Night/night-club doorway	Young men, teens/20s. 'A' pushes 'B' and goes back into club. Another man, 'D', joins group, talks to 'C'.	Man 'C' turns towards 'B'.	'C' assaults 'B'.
10 Night/outside nightclub	Crowd breaks up into smaller groups. Young woman approaches small group of men.	She steps forward raising her arm.	She assaults one of the men.
11 Night/car park	Casually dressed young man on kerb next to parked car and bushes. Walks up and down, looks around.	Turns to step off kerb.	Steals hub-cap.
12 Day/doorway	Scruffy middle-aged man stands leaning against fire hose.	Turns towards glass door.	Kicks door.
13 Night/street	Number of small groups. Three males late teens/early 20s step out onto road, one walking backwards.	'A' holding up hands, walks backwards, in front of 'B'.	'A' lunges at throat of 'B'.
14 Day/car park	Two males, late teens, walking in through car park.	Standing next to passenger door of parked car.	Steal car.

Appendix (continued)

Time/place	Matches		
	event description	scene at pause	after pause
1 Night/street	Mixed groups walking; woman talking to man in group behind her; she steps back, towards him.	Next to man, arm raised.	Continue talking and walking.
2 Night/empty street	Two young men, one runs off pavement towards traffic bollard.	Young man approaching bollard.	Keeps running past bollard.
3 Night/empty street	Man, 30ish, walking along street.	Reaches rubbish bin.	Keeps walking.
4 Night/street	Mixed group of young men; two figures step out of group.	One leans into the other raising arm.	Horseplay.
5 Day/car park	Full 15 s, lone man stands next to two cars.	Starts to move.	Joined by another man who displays parking ticket.
6 Night/outside take-away doorway	Mixed group, teens/20s; female, approaches with arms pointing at another in a small group.	Reaches her, arm raised.	Puts a hand towards her head; embraces her.
7 Night/street	Man walking on street, reaches pub window.	Turns into towards window.	Looks through window, walks away.
8 Day/car park	Man standing between two cars, looking around.	Turns to face door of car behind him.	Opens passenger door and sits in car.
9 Night/street	Five young men, 20s, 'A' and 'B' in animated conversation. 'A' turns away, 'C' turns to follow, 'B' taps on 'C's chest. 'A' turns back, points at 'B'.	'B's head tilted up and mouth wide open.	'A', 'C', and 'D' walk away; 'B' laughing.
10 Night/outside nightclub	Few, mixed group, leave nightclub; woman leaves group of men.	Raises her arm.	Taps man on shoulder, then chats to him.
11 Night/car park	Two casually dressed young men walking on kerb next to parked cars and bushes.	Look around.	Move off, away from car park.
12 Day/doorway	Scruffy middle-aged man standing in doorway.	Leans towards window.	Walks away.
13 Night/street	Small groups, late teens/20ish; one small group standing in road.	Male, puts hands on shoulders of another from behind.	Horseplay.
14 Day/car park	Two males, late teens, walking through car park.	Walking close to cars.	Keep walking and leave car park.

Appendix (continued)

Time/place	Incidents		
	event description	scene at pause	after pause
15 Night/street outside club	Two males, teens/early 20s, 'A' and 'B' walk then stop. 'A' steps out of view. A third male, 'C', arrives, stands next to a young woman, turns his back to 'B', 'A' reappears behind 'C'.	'C' raises his arm.	'A' punches 'B's head.
16 Night/street	Club/pub closing time. Number of groups. In one group of young men, man 'A' shoves 'B'. Man 'C', behind 'A'.	Man 'C' turns to face 'A's back.	Man 'C' assaults 'A' and pulls him to the ground.
17 Night/empty street	Man and woman 30ish walking, talking.	She turns inwards.	She pulls out a 'For Sale' sign.
18 Night/street	Group of males, 20s, widely spaced, running. At pavement 'A' and 'B' hold onto each other.	'C' runs up to join them.	'C' assaults 'A'.

Appendix (continued)

Time/place	Matches		
	event description	scene at pause	after pause
15 Night/street outside pub	Three men and one woman, 20s, on pavement. One man, 'A', leaves, steps out of view.	'A' reappears behind man 'B'.	All four stand chatting.
16 Night/street	Number of small groups. In one group of three someone is on the ground.	Man 'A' reaches down to 'B'.	'A' helps 'B' to stand.
17 Night/empty street	Man and woman 30ish walking, talking.	Turn to face each other.	Continue walking.
18 Night/street	Group of males, widely spaced. Man 'A' runs from pavement to 'B', in middle of street.	'A' puts his arm across body of 'B'.	'A' embraces 'B' and they walk back to pavement.

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 33 2004

www.perceptionweb.com