Exogenous social identity cues differentially affect the dynamic tracking of individual target faces

R. Allen* and F. Gabbert[#]

* School of Psychology, University of Aberdeen

[#] Division of Psychology, University of Abertay

Abstract

We report on an experiment to investigate the top-down effect of exogenous social identity cues on a multiple-identity tracking (MIT) task, a paradigm well-suited to investigate the processes of binding identity to spatial locations. Here we simulated an eyewitness event in which dynamic targets, all to be tracked with equal effort, were identified from amongst a “crowd” of eight faces, as an assailant, bystander, policeman and victim. Even in such a simplistic paradigm, where no actual assault was witnessed and no consequences were associated with the task, results demonstrated a significant attentional bias, namely that participants were significantly better at tracking the assailant, bystander and policeman than they were the victim. We found no difference in accurate recall based upon the use of text or face cues and no systematic pattern of response errors.

Introduction

In 1980, UK police shot and killed a pregnant girl by mistake whilst trying to capture an armed robber who, in a dark alley, used her as a human shield as he fired a shotgun. In 1983, three UK detectives opened fire on a car in which they believed was an armed robber who had escaped from custody. Eight of fifteen bullets fired struck the man before one officer, believing he might continue to fight, closed in and struck the man a heavy blow with the butt of his revolver, fracturing his skull. It was an innocent bystander, not the man they were after. In 1985, in two separate incidents, UK police shot and killed a five-year-old boy whilst searching for his father, believed to be an armed robber, and permanently crippled the mother of another suspected armed robber they were searching for (Rogers, 2003). And in 2005 Jean Charles de Menezes, a Brazilian National, was followed into an underground railway station where anti-terrorist police, falsely believing him to be a suicide bomber, shot him eight times with hollow point (dumdum) bullets, seven times in the head and once in the shoulder (Kennison & Loumansky, 2007).

Whilst such fatal “mistaken identity” shootings are rare, cases in which innocent people are imprisoned and subsequently executed due to mistaken identity are more common. For example, in the US between 1967 and 2002, 32% of innocent individuals in potential capital cases were convicted partly or wholly due to erroneous identification testimony (Koosed, 2009), and just in terms of selecting a suspect from a line-up, eye-witnesses are accurate only 40% of the time, with another 20% of their choices being innocent foils and 40% not producing an identification (Wright & McDaid, 1996).

The increasing use of surveillance officers and CCTV operators also offers opportunities for mistaken identity. Their task involves keeping track of relevant targets (thieves, assailants, colleagues, etc.) within a context that might include many distractors (e.g., from among the general public – see Jean Charles de Menezes above). Covert surveillance officers often work in teams to perform this task at ground level, while CCTV operators track targets remotely by viewing multiple screens within a control room. Despite the importance of good observation and tracking skills in these jobs (recruitment procedures for surveillance officers sometimes require candidates to prove their observation skills prior to interview; see www.mi5.gov.uk/careers/mobile-surveillance.aspx), minimal research exists regarding how such tracking is accomplished. Anecdotal evidence simply suggests that CCTV operators are highly skilled in tracking suspects, even across multiple camera sites and viewing monitors (Scott-Brown & Cronin, 2007). This said, it seems reasonable to assume that the task essentially involves a labelling component (who/what are the targets), a visual component (what do the targets look like) and a spatial component (where are they), which will change as the targets move through the environment and require some attentional separation of targets from distractors. Given the continued frequency of mistaken identity incidents, together with the associated injustice and emotional turmoil, there is a pressing need to uncover the limitations of the requisite perceptual processes, especially those associated with face recognition in a dynamic setting.

One fruitful line of enquiry would be to understand how we bind identity to spatial location, and maintain such bindings over time (i.e., knowing who is where and when). Research has shown that, unlike static tasks, features are not properly conjoined during attentive, dynamic tracking. Motion makes feature binding more difficult to retain, probably because the basis for binding shared location (Treisman, 1988) is no longer a perceptually-stable property (Fahle & Koch, 1995), perhaps because of stronger bindings to earlier locations (Hollingworth & Rasmussen, 2010). To this end a potentially useful experimental paradigm would be some form of the multiple-object tracking (MOT) task (Pylyshyn & Storm, 1988). Although, by minimising the contribution of top-down processes, the MOT task was initially used to examine the low-level, bottom-up processes associated with dynamic spatial cognition it has, more recently, been revised as a multiple-identity tracking (MIT) task to investigate the tracking of individualised targets, such as cartoon animals (Horowitz et al., 2007), butterflies (Botterill, Allen & McGeorge, 2011) and unfamiliar faces (Ren, Chen, Liu & Fu 2009). This task appears to make greater demands on top-down attentional processes. Successful performance requires that we maintain bindings of identities to changing spatial locations.

Botterill et al. (2011), for example, confirmed Horowitz et al.'s (2007) observation that MIT performance was only half that of what one might expect from a similar MOT task, i.e., 2 identities accurately located as compared to 4 with the MOT task. Further, when they analysed participants' incorrect responses, they found that, even when participants were wrong about where a particular target object was located, in ~90% of cases the response they gave still came from the target set, suggesting that <u>location</u> tracking remained as high as ever. This finding has been interpreted as being indicative of two separate processes, one of maintaining spatial locations, the other of binding identities to maintained spatial locations (but see Cohen, Pinto, Howe & Horowitz, 2011). In particular, Oksama and Hyönä (2008), in their model of multiple-identity tracking (MOMIT) theory, have suggested that whereas the creation, storage and, presumably, updating of identity/location bindings occurs in working memory's episodic buffer, spatial location coordinates are held in visual short-term memory (VSTM); though, they do not appear to make explicit the significant performance differences between MOT and MIT within the model.

One of the dominant theories about the integration of object-specific information is that proposed by Kahneman, Treisman and Gibbs (1992), who described the concept of an "object file", a cognitive structure into which featural information can be saved. They showed, in a target-naming task, that performance improved if a preceding random display had also contained the target, and that performance was better still if the preceding display contained the same target in the same location. Kahneman et al. interpreted the latter effect, what they called an "*object-specific preview benefit*", as resulting from the repetition of a conjunction of features (i.e., target and location) already stored in an object file. Subsequently, van Dam and Hommel (2010) have suggested that the binding of features into a single object file may not be limited to those that are task-relevant. Task-irrelevant features, such as text descriptors, facial images and updating spatial locations, may also be bound into the same object file, whether or not they are part of the same object, just so long as they overlap spatially.

It may, however, be that identity tracking performance is not consistent across object types. After all, faces, for example, are seen as being special objects that are preferentially processed. Oksama and Hyönä (2008) were the first to use faces in an MIT experiment. However, they were interested in how facial familiarity affected identity tracking performance, reporting that familiar faces were more easily tracked than pseudo-faces.

A more recent study by Ren et al.(2009), over 5 experiments, investigated the perceptual characteristics of identifying and tracking unfamiliar faces. They found that the identification of target faces (not their spatial locations) was poorer when they were displayed dynamically, compared to statically, although performance was always better than chance. They too confirmed that identity tracking was significantly poorer than location tracking and also provided evidence that target faces are, at least partially, processed implicitly, even when not task-relevant, and this impairs tracking performance (e.g., when tracking different upright target faces). Such findings have come closer to realistic settings, such as the tracking of individual faces in a crowd, but have not yet incorporated social meaning, such as who or what each individual face might represent or the nature of any relationship they might have with each other or the observer.

This is pertinent in light of recent research by Howard, Gilchrist, Troscianko, Behera and Hogg (2009), who found that high-level target characteristics, such as semantic evaluation, predicted visual search patterns in a realistic scene significantly more so than low-level characteristics such as appearance or salience. Indeed, MIT in real-life is rarely as simple as tracking non-individuated targets in a semantic void. Instead, contexts typically have some level of personal or semantic relevance, for example, a mother keeping track of her children in a park.

The literature on social cognition is informative with regard to context. Frith and Frith (2006), for example, talk about the interplay between implicit and explicit knowledge. They suggest that whilst stereotypical prejudice can be triggered automatically, this can be controlled by the top-down control of explicit knowledge. This may be particularly so for the formation of evaluative associations with faces (Todorov & Olson, 2008), or affective person knowledge, whose acquisition comes from minimal, in the former case behavioural, information. Such information is spontaneously retrieved at face perception, activating neural systems for social cognition and emotion analysis (Todorov, Gobbini, Evans & Haxby, 2007). These ideas are encapsulated in Bar (2007) who claims that our memories are as important as our sensory experiences to the interpretation of our environment. His proposal is that sensory input triggers analogous representations in memory. Such co-activations, of past experience, provide moment-by-moment predictions of the most likely associations to the specific situation. Such constantly updating predictions are intended to guide our thoughts, plans and actions.

A number of face-recognition studies have highlighted the advantage of a semantic or holistic decision-making process, as compared to a feature-based one (e.g., Olsson & Juslin, 1999). One reason for this is that trait judgements are deeper and therefore more elaborate, leading to better memory performance (e.g., Coin & Tiberghien, 1997; Gallo, Meadow, Johnson & Foster, 2008).

An additional issue, particularly in terms of the eyewitness literature, concerns the effect of emotion upon subsequent memory. Whereas some studies have suggested that negative-emotion events (stress/violence) impair eyewitness memory (e.g., Loftus & Burns, 1982; Peters, 1988) others suggest such memories are well retained (e.g., Bohannon, 1992; Heuer & Reisberg, 1990). Others still, such as Christianson (1992), have supposed that negatively-charged events attract preferential attention from early perceptual processing associated with arousal, distinctiveness etc. and later conceptual processing such as elaboration, which result in the better retention of central as compared to peripheral detailed information.

It seems clear, then, that top-down processes have the capacity to impact upon tracking performance. But the degree, type and homogeneity of such an attentional bias seem, as yet, uncertain.

The central question for this research, then, is to what extent social labels impact upon memory, in terms of tracking performance? We therefore propose to carry out a two-part MIT task in which the stimuli are established in a social setting, namely a “crowd” represented by a group of faces, in which an assault has previously taken place. Certain individuals, identified as targets to be tracked, will be textually labelled as an assailant, a bystander who might be either a confederate of the assailant or an eye-witness, a policeman and a victim. Because these are, typically, the characters you expect to find at this kind of “event”. Such a paradigm will allow us to quantify the extent and distribution of any top-down effect by investigating, in general, whether a social setting impacts upon location and/or identity tracking performance.

This paradigm also provides for a number of opportunistic, subsidiary analyses. Since successful tracking requires both facial and textual information be bound to changing spatial locations, facial and textual cues will be used to probe performance. This will provide an opportunity to evaluate which better promotes accurate recall, text cues (as used by Horowitz et al., 2007) or image cues (as used by Botterill et al., 2011). Further, by presenting blocks of single-cue and mixed-cue trials we can examine the extent to which the retrieval processes cued differ, in terms of performance. By always presenting the less demanding single-cue trials first, thus minimising differences in performance, we can also be more confident that any significant difference found is more likely to be a processing, rather than levels of difficulty, difference. Finally, though not specifically designed to address this issue rigorously, we can examine whether participants show any racial bias towards any of the experiment's social roles or whether, given any probe, an incorrect response is more frequently associated with one role rather than another.

Method

Participants:

There were 48 participants (13 males) aged 17 – 29 years (M = 20.54, SD = 2.20), of whom 44 were Caucasian. All had normal or corrected-to-normal vision. They were all naïve as to the purpose of the experiment and gave informed consent.

Materials:

Each trial contained the same 8 stimuli in the form of images of different, clean-shaven young male faces (200+ colours), oval-cropped to exclude everything but the eyes and eye-brows, nose and mouth, on a white background, presented on a 19" computer monitor (viewing area ~37° of horizontal x ~30° of vertical visual angle, viewed from a distance of ~57cm; screen resolution 1280 x 1024), using the Superlab 4.0.4 (Cedrus Corporation) programme running on a 60Hz pc. Two each of the faces were Caucasian, Afro-American, Asian and Arabic (see figure 1). Only 8 faces were chosen because the intention was to study social labels rather than race or face-specific effects. In addition, as this is the first study of its kind, we wanted to investigate the general case of a heterogeneous, rather than homogenous, crowd. Each face subtended ~3.5° of vertical and ~2.5° of horizontal visual angle. Each trial consisted of a static (Target Acquisition (TA)) phase, a dynamic (Target Tracking (TT)) phase and a Retrieval phase. The TA phase lasted ~3s and, during this period, 4 faces were pseudo-randomly identified as targets. This was done by each target face being circled in red. These targets were also pseudo-randomly labelled to identify their "role" in the trial (i.e., assailant, bystander, policeman or victim). Text labels were placed adjacent to the relevant image. At the start of the TT phase, the red circles and the text labels disappeared, and all 8 faces moved randomly at ~6.8°/s for ~6s. Note that only the type of cue used distinguished the text-cue from face-cue trials (i.e., their object trajectories throughout the TT phase were identical). At the end of this tracking period all 8 faces were obscured by red discs numbered 1 – 8 (Retrieval phase). This numbering was also pseudo-randomised so that there was no consistent link between a particular face and number. Thereafter, a series of four retrieval cues were displayed, one at a time, to test for each target. These were either in the form of a text probe (e.g., "Where is the assailant?") or a face cue (e.g., "Where is . . . .") in which the image of a particular target face was inserted.

Design:

This was a forced-choice experiment consisting of 3 blocks, each of 40 trials. One block consisted of all text cues (20 different trials repeated twice) another of all face cues (again 20 different trials repeated twice) with the last being a mix of text and face cues trials (i.e., the 20 text-cue trials plus the 20 face-cue trials), all randomly presented. In each of the 20-trial sets, each of the race faces was one of the four target characters five times. At times this meant that two of the same race faces were targets at the same time, but this applied equally to all four of the race faces. Further, since the order in which the trials were presented was randomised by the software, it was possible for faces from the same race to be identified as the same character sequentially. More importantly, however, exactly the same animations were used for both the face-cue and text-cue trials, so there was absolutely no difference in tracking workload.

Whether participants did the block of text-cue trials and then the mixed block of trials (part 1), or the block of face-cue trials followed by the mixed block of trials (part 2) was pseudo-randomised. The mixed block of trials was always completed last to avoid a workload imbalance, given how much harder the mixed block task was compared with the blocks of only text-cue or face-cue trials.

Procedure:

Participants were tested individually in a small, window-less room. Instructions were displayed on the monitor and given verbally. They were told that the experiment involved a crowd of people, represented by their faces. In each trial 4 of the 8 individual faces, labelled as an Assailant, his Victim, a Bystander and a Policeman, would be their targets. The bystander might be a witness to the assault or could be a confederate of the assailant - to reproduce the uncertainty that is often the case in such "eye-witness" situations. When the identifying labels disappeared all the faces would begin to move randomly around the screen. Their task was to track the location of each of the target individuals. After a period of motion, the faces would stop moving and be completely covered over by numbered red discs. Thereafter, they would be probed as to the location of each of the tracked individuals, one at a time. In the block of text-cue trials it would be of the form "Where is the Victim?", whereas in the face-cue trials it would be of the form "Where is . . . ", with the appropriate face inserted. In the mixed block, all of the text-cue and face-cue trials were presented again in a random order so that participants could not anticipate the probe format. The order of the text/face cues was counterbalanced. Participants recorded their response using the number keys above the QWERTY keys of the computer keyboard.

Results

Several data sets were extracted from the findings that were then arc sine transformed before being analysed (for a summary of the original % correct tracking performances see table 1). Note, in all analyses, an alpha value of .05 was used, wherever necessary the Greenhouse-Geisser correction was applied, and all pair-wise comparisons used the Bonferonni correction for multiple comparisons. Chance levels for an identity tracking task trial vary, typically, between 12.5 and 20%, provided that previous responses have been accurate.

In order to answer the central question of to what extent social labels impact upon memory, in terms of tracking performance, the first set extracted was the identity tracking performance (i.e., how accurately participants could relate each specific target to its final location). The next set, much as Botterill et al. (2011), was the location tracking performance (i.e., how frequently participants' responses were chosen from members of the target subset, irrespective of whether or not they had managed to locate the specific target correctly). A separate analysis of location tracking performance may, additionally, provide evidence of whether identity and location tracking share a common process or are separable. This is particularly so for the analysis of Character where, for a shared process, you would expect no differences in any tracking bias between identity and location performance. Since identity tracking is a subset of location tracking, these two data sets were analysed separately.

Identity tracking performance

A repeated-measures ANOVA was carried out with Character (Assailant, Bystander, Policeman, Victim) and Mix (single-cue trials, mixed-cue trials) as the within-subjects factors, and Cue (text, face) as the between-subjects factor.

There was a significant main effect of Character ($F(3, 138) = 8.20$, $p < .001$, $\eta_p^2 = .15$). Pair-wise comparisons showed that the Victim (M = .617, SD = .16) was successfully tracked significantly less frequently than the Assailant (M = .691, SD = .19, p < .001) and the Bystander (M = .672, SD = .17, p = .019) (see figure 2). No other pair-wise comparisons were significant.

There was also a significant main effect of Mix ($F(1, 46) = 11.45$, $p = .001$, $\eta_p^2 = .20$), with performance on the single-cue trials (M = .695, SD = .17) being significantly higher than that on the mixed-cue trials (M = .620, SD = .19). No other main effects or any interactions were significant. In particular, that there was no main effect of Cue, or any significant interaction with it, demonstrates that text and face probes did not differ in their ability to trigger memory recall.

Location tracking performance

A repeated-measures ANOVA was carried out with Character (Assailant, Bystander, Policeman, Victim) and Mix (single-cue trials, mixed-cue trials) as the within-subjects factor and Cue (text, face) as the between-subjects factor.

There was a significant main effect of Character ($F(3, 138) = 8.66$, $p < .001$, $\eta_p^2 = .16$). Pair-wise comparisons showed that the Victim (M = .965, SD = .11) was successfully tracked significantly less often than both the Assailant (M = 1.034, SD = .12, p < .001) and the Policeman (M = 1.003, SD = .15, p = .046). Additionally, the Bystander (M = .994, SD = .15) was also tracked significantly less frequently than the Assailant (p = .040) (see figure 3). No other pair-wise comparisons were significant.

There was also a significant main effect of Mix ($F(1, 46) = 6.51$, $p = .014$, $\eta_p^2 = .12$), with performance on the single-cue trials (M = 1.013, SD = .12) being significantly higher than that on the mixed-cue trials (M = .985, SD = .14). No other main effects or any interactions were significant. Again, that there was no main effect of Cue, or any significant interaction with it, demonstrates that there was no advantage to using text or face probes for memory recall.

Finally, we addressed the secondary questions raised in the introduction. Since there were only 4 non-Caucasian participants, no analysis of bias amongst them was possible. But we could examine whether participants showed a race or role bias across the stimuli, or whether there were systematic errors across the characters. As to the former, repeated-measures ANOVAs were carried out with Race (Caucasian, Afro-American, Asian, Arabic) and Character (Assailant, Bystander, Policeman, Victim) as the within-subjects factors for identity and location performance separately. Note that here results were collapsed across Cue (text/face) and Mix (single-cue trials, mixed-cue trials).

For identity tracking, there was a significant main effect of Race ($F(3, 141) = 9.60$, $p < .001$, $\eta_p^2 = .17$). Pair-wise comparisons showed that Asian faces were successfully tracked significantly less often than any of the other races (all p's < .05). There was also a significant main effect of Character (F(3, 141) = 16.46, p < .001, η_p^2 = .26). Here, pair-wise comparisons revealed that Victims were successfully tracked significantly less often than any of the other characters (all p's < .001). These effects were moderated by a significant interaction between them (F(9, 423) = 16.17, p < .001, η_p^2 = .26) (see figure 4). Pair-wise comparisons showed that, a Caucasian Assailant was successfully tracked significantly more often than any other Assailant (all p's < .05). Caucasian and Afro-American Bystanders were successfully tracked significantly more often than Arab or Asian Bystanders (p's < .05), but did not differ themselves (p = .515). Further, Asian Bystanders were the least successfully tracked (all p's < .01). Caucasian Policeman were the least successfully tracked (all p's < .001), there being no other significant differences in performance. Finally, Asian Victims were the least successfully tracked (all p's < .01), there being no other significant differences in performance.

For location tracking, there was again a significant main effect of Race (F(2.45, 115.24) = 9.73, p < .001, η_p^2 = .17). Pair-wise comparisons showed that Caucasian faces were successfully tracked significantly more often than any of the other races (all p's < .01). There was also a significant main effect of Character (F(3, 141) = 17.20, p < .001, η_p^2 = .27). Here, pair-wise comparisons revealed that Victims were successfully tracked significantly less often than any of the other characters (all p's < .01). In addition, the Bystander was also tracked significantly less often than the Assailant (p = .04). However, these effects were moderated by a significant interaction between them (F(6.41, 301.08) = 19.64, p < .001, η_p^2 = .29) (see figure 5). Pair-wise comparisons showed that, a Caucasian Assailant was successfully tracked significantly more often than any other Assailant (all p's < .01). Caucasian and Afro-American Bystanders were successfully tracked significantly more often than Arab or Asian Bystanders (p's < .01), but did not differ themselves (p = 1.000). Further, Asian Bystanders were the least successfully tracked (all p's < .01). For the Policeman, there were no significant differences in performance across races and, finally, Arab and Caucasian Victims were the more successfully tracked (both p's < .01), although they did not differ significantly from each other (p = .054).

Lastly, to discover whether an incorrect response tended to be more commonly of one role rather than another (i.e., that there were systematic errors amongst the characters), four repeated-measures ANOVAs of the identity tracking data were carried out, with Error Choice (i.e., 3 from Assailant, Bystander, Policeman, Victim) as the within-subjects factor.

For Assailant probes there was a significant main effect of Error Choice (F(2, 94) = 9.50, p < .001, η_p^2 = .17). Pair-wise comparisons showed Victim (M = .451, SD = .06) was the more likely to be chosen in error than either Bystander (M = .410, SD = .07; p = .008) or Policeman (M = .399, SD = .08; p < .001).

For Bystander probes there was a significant main effect of Error Choice (F(2, 94) = 7.97, p = .001, η_p^2 = .15). Pair-wise comparisons showed Policeman (M = .477, SD = .08) was the more likely to be chosen in error than either Assailant (M = .412, SD = .07; p = .001) or Victim (M = .429, SD = .09; p = .012).

For Policeman probes there was a significant main effect of Error Choice (F(1.67, 78.43) = 3.56, p = .041, η_p^2 = .07). However, pair-wise comparisons revealed no significant differences between the three choices: Assailant (M = .444, SD = .07), Bystander (M = .409, SD = .08) and Victim (M = .450, SD = .08).

Finally, for Victim probes, there was no significant main effect of Error Choice ($F(2, 94) = 0.02$, $p = .976$), with Assailant (M = .425, SD = .08), Bystander (M = .427, SD = .06) and Policeman (M = .427, SD = .06) chosen equally as often (see figure 6).

Discussion

The study's primary goal was to investigate whether a social setting would impact upon identity and location tracking performances and, in particular, to characterise the nature of any attentional bias. In the Target Acquisition phase of the current study, participants had three seconds to associate four unfamiliar faces with one of four labels (Assailant, Bystander, Policeman or Victim), and then spatially track the targets whilst they moved. The face/label pairings were randomised for each trial. Thus, faces could be uniquely identified by their appearance and their label, but only the label contained meaningful/semantic information. Results for both identity and location tracking were very similar, namely that there were significant differences in tracking performance, dependent upon the character of the target face, with victims being tracked significantly less well than other characters.

Even though this paradigm was simplistic, with all the faces being randomly assigned to fulfil each of the social roles and no acts or threats of violence being portrayed, this result suggests that participants implicitly evaluated the targets' roles, because to do so explicitly would be to contradict their tracking instruction, and determined that the victim merited less attention than did the other characters, or that the other characters merited more attention, to the detriment of tracking the victim. Whilst not a real event, it's reasonable to speculate that the post-crime scene presented in the experiment would trigger many long-term memories of similar events, either experienced first-hand or vicariously on the news or in television programmes. According to Mather and Sutherland (2011, p.114) " . . . [a] wide range of cognitive and emotional challenges increase autonomic arousal ". They give the example of an emotional picture presented for a few seconds but affect words, too, can have a similar effect (e.g., Kensinger & Schacter, 2006). Further, Mackay et al. (2004) claim that arousing words receive superior bindings by triggering emotional reactions via the amygdala, which then triggers the hippocampus to activate attention to bind the arousing stimuli to other contextual features, thus enabling superior encoding and retrieval. This is much in keeping with Bar (2007) which suggested that analogous representations of sensory input, based on past experience, are triggered to guide our thoughts, plans and actions via a process of constantly-updating predictions of what typically follows as a consequence.

In this case, one might speculate that past experience would suggest that, in a real situation involving an assailant, a bystander who might be the assailant's confederate, a policeman and a victim, one is typically likely to be at less risk, or expect to receive less help from the victim than the other three characters. Thus, one need be less alert to the victim's whereabouts. Alternatively, such an attentional bias has been discussed in terms of an inability to disengage from a threat, at least from the assailant and bystander (Fox, Russo & Dutton, 2002). Finally, note that, whilst there were some differences between identity and location tracking performance of individual characters, these findings are inconclusive. Further studies are required to establish conclusively whether identity and location tracking consist of a single process or not.

As to the secondary questions raised in the introduction, that there was no main effect of or significant interaction with the type of cue, text or face, suggests that neither is better at promoting accurate recall. However, that performance on a block of single cue trials, either text or face, was significantly better than that on a block of mixed text and face cue trials suggests a divergence in the necessary processing for each, such that having to task switch leads to a reduction in performance. In fact, faces have, previously, been shown to activate a functionally defined face-responsive region in the right fusiform gyrus and words, a word-responsive region in the left inferior temporal gyrus, and such activations were larger for trials with emotional content (Hofstetter, Achaibou & Vuilleumier, 2012), thus implying a qualitative difference between the two.

The identity and location tracking data were also examined to compare performance on each character, by race, to investigate whether participants showed a racial bias towards any of the experiment's social roles. Results showed that, for both identity and location tracking, for Caucasian faces performance was best when they represented the assailant or bystander; for Afro-American faces, when they represented the bystander; for Asian faces, when they represented the assailant or the policeman; and, for Arabic faces, there was no difference in performance whatever they represented.

Finally, to investigate whether there was a pattern of incorrect responses, the distribution of response errors amongst the identity tracking performance data was examined. Results showed that, for assailant probes, participants' most likely error response was victim; for bystander probes, the most likely error response was policeman; and for policeman and victim probes, error responses were equally distributed across the other three alternatives. There seems, therefore, to be no systematic distribution of response errors.

The findings in this experiment have relevance for understanding the nature of spatial/identity bindings in real life. As mentioned in the introduction, some jobs specifically require these skills. A recent Home Office Report outlines the need to apply findings from psychological science to improve the CCTV review process (Hillstrom, Hope & Nee, 2008). The report calls for a clearer understanding of psychological factors influencing the effectiveness of CCTV operators, and recommends a Cognitive Task Analysis of the process to identify particular areas of difficulty or limitations. The present study simulates a task of a CCTV operator, albeit in a more controlled and artificial way. Nevertheless, it is an initial step towards understanding how people cope with dynamic visual and semantic information in the real world.

Acknowledgements

Thanks to Dr Todd Horowitz and four other unnamed reviewers for their valuable comments.

References

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11 (7)*, 280-289.

Bohannon, J. N. (1992). Arousal and memory: Quantity and consistency over the years. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: The problem of "flashbulb" memories* (pp. 65-91). New York: Cambridge University Press.

Botterill, K., Allen, R. & McGeorge, P. (2011). Multiple-object tracking: The binding of spatial location and featural identity. *Experimental Psychology, 58(3)*, 196-200.

Christianson, S-Å. (1992). Emotional Stress and Eyewitness Memory: A Critical Review. *Psychological Bulletin, 112 (2)*, 284-309.

Cohen, M. A., Pinto, Y., Howe, P. D. L. & Horowitz, T. S. (2011. The what–where trade-off in multiple-identity tracking. *Attention, Perception & Psychophysics, 73*, 1422–1434. DOI 10.3758/s13414-011-0089-7

Coin, C. & Tiberghian, G. (1997). Encoding activity and face recognition. *Memory, 5 (5)*, 545-568.

Fahle, M. & Koch, C. (1995). Spatial displacement, but not temporal asynchrony, destroys figural binding. Vision Research, 35, 491-494.

Fox, E., Russo, R. & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cognition & Emotion, 16 (3)*, 355-379.

Frith, C. D. & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*, 36-46.

Gallo, D. A., Meadow, N. G., Johnson, E.L. & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58 (4)*, 1095-1111.

Heuer, F. & Reisberg, D. (1990). Emotion, arousal and memory for detail. In S-Å. Christianson(Ed.), *The handbook of emotion and memory: Research and theory (pp.* 151-180). Hillsdale, NJ: Erlbaum.

Hillstrom, A., Hope, L. & Nee, C. (2008) *Applying psychological science to the CCTV review process: a review of cognitive and ergonomic literature.* Project Report. HMSO, London.

Hofstetter, C., Achaibou, A. & Vuilleumier, P. (2012). Reactivation of visual cortex during memory retrieval: content specificity and emotional modulation. *Neuroimage, 60 (3)*, 1734-1745.

Hollingworth, A. & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. Journal of Experimental Psychology: Human Perception & Performance, 36, 543-564.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A. & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69 (2)*, 172-184.

Howard, C. J., Troscianko, T., Gilchrist, I. D., Behera, A., & Hogg, D. C. (2009). Searching for threat: factors determining performance during CCTV monitoring. In D. de Waard, J. Godthelp, F.L. Kooi & K.A. Brookhuis (Eds.) *Human Factors, Security and Safety* (pp. 1 - 7). Maastricht, the Netherlands: Shaker Publishing.

http://www.stockwellinquest.org.uk

Kahneman, D., Treisman, A. & Gibb, B. J. (1992). The reviewing of object files—object-specific integration of information. *Cognitive Psychology, 24(2)*, 175 – 219.

Kennison, P. & Loumansky, A. (2007). Shoot to kill – understanding police use of force in combating suicide terrorism. *Crime, law and social change, 47*, 151–168.

Kensinger, E. A. & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience, 6*, 110-126.

Koosed, M. (2009). Reforming eyewitness identification law and practices to promote the innocent. *Creighton Law Review*, *595*, 1-56.

Loftus, E. F. & Burns, T. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition, 10,* 318-323.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L. & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition, 32*, 474-488.

Mather, M. & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*, 114-133.

Oksama, L. & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56*, 237–283.

Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *Journal of Applied Psychology, 84*, 42-49.

Peters, D. P. (1988). Eyewitness memory and arousal in a natural setting. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes(Eds.), *Proceedings of the Second International Conference on Practical Aspects of Memory (Vol.* 1, pp. 89-94). New York: Wiley.

Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197.

Ren, D., Chen, W., Liu, C. H. & Fu, X. (2009). Identity processing in multiple-face tracking. *Journal of Vision, 9 (5)*, 1–15.

Rogers, M. D. (2003). Police force! An examination of the use of force, firearms and less-lethal weapons by British police. *The Police Journal, 76*, 189-203.

Scott-Brown, K. C. & Cronin, P. D. J. (2007). An instinct for detection: psychological perspectives on CCTV surveillance. *The Police Journal, 80*, 287-305.

Todorov, A., Gobbini, M. I., Evans, K. K. & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia, 45*, 163-173.

Todorov, A. & Olson, I. R. (2008). Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdale and temporal pole are damaged. *Social Cognitive and Affective Neuroscience, 3 (3)*, 195-203.

Treisman, A. (1988). Features and objects, the Fourteenth Bartlett Memorial Lecture. Quarterly Journal of Experimental Psychology A, 40, 201-237.

van Dam, W. O. & Hommel, B. (2010). How object-specific are object files? Evidence for Integration by Location. *Journal of Experimental Psychology: Human Perception and Performance, 36 (5)*, 1184–1192.

Wright, D. B. & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology, 10*, 75-84.

Tables

Table 1: original % correct, by character and trial type, across identity and location tracking, collapsed over probe type

Task	blocked trials			
	assailant	bystander	policeman	victim
	M (SD)	M (SD)	M (SD)	M (SD)
identity tracking	47.03 (16.09)	44.95 (15.24)	45.34 (14.87)	39.87 (15.23)
location tracking	74.66 (10.18)	70.91 (11.91)	73.57 (10.89)	69.14 (10.43)

Task	mixed trials			
	assailant	bystander	policeman	victim
	M (SD)	M (SD)	M (SD)	M (SD)
identity tracking	38.75 (16.85)	37.14 (15.74)	35.78 (15.17)	33.02 (14.31)
location tracking	71.87 (11.99)	69.06 (12.89)	69.01 (12.61)	65.57 (11.41)

Figures

Figure 1: face stimuli

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6