Unblinking Eyes: The Ethics of Automating Surveillance

Abstract

In this paper I critique the ethical implications of automating CCTV surveillance. I consider three modes of CCTV with respect to automation: manual (or non-automated), fully automated, and partially-automated. In each of these I examine concerns posed by processing capacity, prejudice towards and profiling of surveilled subjects, and false positives and false negatives. While it might seem as if fully-automated surveillance is an improvement over the manual alternative in these areas, I demonstrate that this is not necessarily the case. In preference to the extremes I argue in favour of partial-automation in which the system integrates a human CCTV operator with some level of automation. To assess the degree to which such a system should be automated I draw on the further issues of privacy and distance. Here I argue that the privacy of the surveilled subject can benefit from automation, while the distance between the surveilled subject and the CCTV operator introduced by automation can have both positive and negative effects. I conclude that in at least the majority of cases more automation is preferable to less within a partially-automated system where this does not impinge on efficacy.

Keywords

Surveillance, automation, CCTV, SUBITO, operator, prejudice, profiling, false positives, false negatives, Clive Norris, Gary Armstrong

Introduction

This paper critiques the ethical implications of automating CCTV surveillance. It is especially concerned with questions of efficacy. Efficacy is a morally significant concept: the likelihood of success has long been seen as an important element in the just war tradition, governing whether a state is justified in declaring war. It has typically been thought that without a chance of success the war would be unjustified. Likewise in surveillance: a system which could not achieve its ends would be unwarranted, gratuitous. A system which could not achieve its ends and which introduced *prima facie* harms (i.e. invasions of privacy in the case of surveillance) would be pernicious. I argue that there are strong concerns with the ability of manually-operated CCTV to produce consistent and reliable results. However, I show that the automation of CCTV risks failing to produce reliable results owing to, often unintentional, embedded prejudice. In preference to either of these extremes, I argue that partially-automated surveillance, in which the automation alerts an operator to potential threats, is the most efficacious system.

Having established the moral preference of partial automation, I then examine the further ethical considerations of privacy and distance. These have an impact on the quantity of information returned to the operator in a partially automated system. Each shall be explained as they are encountered, and each shall be considered from two perspectives of more or less information being returned. I argue that partial automation is not free from ethical concerns, and even introduces some of its own. However I show that none of these is over-riding of the considerations of efficacy presented earlier. Furthermore, I argue that in at least the majority of cases more automation is preferable to less within a partially-automated system where this does not impinge on efficacy.

1 Manual Surveillance

CCTV has a number of benefits over other forms of surveillance. It has a broad visual scope with the potential to cover a 360º area simultaneously; the camera itself is not selective in whom it watches; and it provides a searchable record which trumps human memory in longevity, authority and accuracy (Gerrard et

al. 2007). Information recorded is available for retrospective data-mining; may be suitable for facial recognition; and can be cross-referenced against other databases in real time. Nonetheless problems persist in using CCTV, often in relation to the operator rather than the system.

This is perhaps not surprising. There are a number of reasons contributing to CCTV operator error. The first of these is the limited processing capacity of the operator. This can lead to the filtering of information by profiling people or people groups, which may itself be based on prejudiced social stereotypes. Taken together, limited processing capacity and prejudicial filtering contribute to an increased likelihood of false positives and false negatives. More false positives and false negatives may result in a system which is less effective at recognizing and responding to security threats. As stated above, a system which is less effective and carries some cost is, *ceteris paribus*, less ethical. I shall therefore consider these three concerns (processing capacity, profiling and prejudice, and false positives/negatives) in turn.

1.1 Processing Capacity

In 2007 Glasgow's 420 CCTV-linked monitors were monitored by eight operators, or 50 per person during the week (McKinnon 2007). This is not atypical. In 2011 Corby Borough Council had two operators monitoring up to 67 cameras at any one time and Worcester City Council had just one employee watching up to 100 monitors (Corby Borough Council Electronic Information Team 2011; Manning 2011). If there is activity on every monitor then this can quickly overwhelm the operator. To avoid this the operators must somehow filter the information. However, this process of filtration often draws on social stereotypes in determining who to target, which risks the operator missing crimes by non-stereotypical offenders (Norris 2002). Operators are also subject to both inattentional blindness and change blindness. Inattentional blindness occurs when an observers' attention is fixed on one particular detail such that they do not notice anomalous events occurring in the same scene (Mack 2003). Change blindness, on the other hand, describes the inability to notice large changes in visual scenes which occur during a momentary distraction such as a blink or the panning of a camera (Simons & Ambinder 2005; see also Resnick 2002). A fourth consideration is operator boredom. For many operators there is little of interest happening most of the time, and so it is hard to maintain the necessary levels of attention for long periods. Taking the phenomena of inattentional and change blindness in addition to information overload and operator boredom, it is not surprising that operator error can sometimes occur.

1.2 Profiling and Prejudice

Given the volume of information the operator faces he must filter it to prioritize his attention. How that filtering takes place is of considerable importance as this will lead to certain people receiving greater attention than others, and so that filtering should be based on criteria that are relevant. In their study of operator behaviour in 1995-96 Clive Norris and Gary Armstrong (1999) argued that in practice the selection criteria used for surveillance were overwhelmingly determined by age, ethnicity and sex. For example, of those surveilled by operators in the course of the study, 39% were teenagers, a proportion which rose to 65% when the operator recorded "no obvious reason" for whom he was watching (i.e. the surveilled subjects exhibited no suspicious behaviour). Yet teenagers comprised less than 15% of the overall UK population, accounted for 23% of cases in which the police were deployed (a total of 10 incidents), and made up just 18% of arrests made on the basis of the surveillance (Norris & Armstrong 1999). It is possible that there may have been a disproportionate number of teenagers in the area under surveillance, and it is not uncommon that teenagers are often let off with a warning rather than arrested and sentenced. Nonetheless, these considerations do not seem to account for the contrast between the proportion of teenagers surveilled relative to the threat they apparently posed (Haggerty 2009).

1.2.1 Group Profiling

Surveillance which allows for the surveillant to indulge his own prejudices makes it likely that some groups will persistently be treated differently (Lyon 2002). It may be that in some cases groups do render themselves liable for surveillance (certain aggressive gangs, for example, who self-identify through sporting unique insignia) in which case filtering the surveillance by group would be appropriate. If on the

other hand the group has done nothing to make itself liable for surveillance then that surveillance would be inappropriate. Judging by the relative number of teenage offenders it could be argued that the operators in the Norris and Armstrong study were therefore guilty of inappropriate surveillance.

1.2.1.1 Stigmatization and Harassment

There are two main concerns with profiling based on membership of a group: unjustified stigmatization and harassment of the innocent. These can be illustrated across four scenarios which, using teenagers and shoplifting as an arbitrary example, contrast the hypothetical situation between every teenager shoplifting, only teenagers shoplifting, some teenagers shoplifting and some non-teenagers shoplifting. The four alternatives are:

i) every teenager and only teenagers engage in shoplifting;
ii) every teenager but not only teenagers engage in shoplifting;
iii) not every teenager but only teenagers engage in shoplifting; and
iv) not every teenager and not only teenagers engage in shoplifting.

For the sake of the illustration I will assume that the surveillance is 100% effective and fully warranted in the case of apprehending shoplifters. I will also assume that the facts regarding teenagers and shoplifting are known in each scenario and that these facts will not change over time. For the time being I will leave aside the question as to how different these cases are from reality.

In scenario (i) it would be unobjectionable to carry out surveillance on teenagers. Given that the surveillance is effective and warranted, then if every teenager shoplifts and no-one who is not a teenager shoplifts profiling would be justified (Lippert-Rasmussen 2010). While it might be argued that this stigmatizes teenagers, it does not do so unjustifiably. They are singled out as shoplifters because they are unique in their shoplifting, but no innocents have been stigmatized.[1]

Scenario (ii) presents a more difficult case: every teenager shoplifts, so focusing attention purely on teenagers would capture all shoplifting teenagers. However not only teenagers shoplift, so older (or younger) shoplifters will go unchallenged. This is perhaps unfair in treating just teenagers as shoplifters when this is not the case. Other members of society also steal from shops and so teenagers are no different in this respect. However the fact remains that all teenagers are shoplifters. The central problem here is how effective such profiling of teenagers would be: it seems important to ask how many non-teenagers shoplift. We might imagine a city, Cleptopolis I (population 10,000) in which there are 2,000 teenagers, all of them shoplifters. However there are 5,000 shoplifters in all. In this case the majority of shoplifters are in fact not teenagers, and so focusing on teenagers would have limited, albeit perhaps significant, impact on reducing shoplifting. To catch the majority of the shoplifters a further policy would need to be put in place, in the absence of which the surveillance would be far from perfect. In neighbouring Cleptopolis II, given the same population and age distribution, the total number of shoplifters is 2,001. There is one adult who couldn't kick the habit as he passed into his 20s. In Cleptopolis II the scenario appears to be very close to scenario (i) in that every and (almost) only teenagers shoplift. As such an exclusive focus on teenagers in Cleptopolis II would be more justifiable than in Cleptopolis I, given that such a focus would miss only one shoplifter in Cleptopolis II whereas it would miss 3,000 (the majority) in Cleptopolis I. However in neither case does it seem as if teenagers are being stigmatized unjustifiably nor are the innocent being harassed.

Scenario (iii) envisages the situation in which only teenagers engage in shoplifting, but not all teenagers are shoplifters. In this case focusing surveillance on teenagers does lead to unjustified stigmatization and harassment of the innocent. The innocent may see themselves as being treated as if guilty, lumped in with the genuinely guilty simply because of their age. They are treated with suspicion owing to the actions of a number of others in their group. They might not know the others or have any influence over their actions. Nonetheless they suffer the consequences of those actions through no fault of their own. It may be argued, however, that it is not teenagers that are being targeted. Rather, we start with the presumption that everyone should be targeted but then those groups *known* to be innocent are *excluded* from targeting. It just

[1] This also does not deny that there might be deep-seated reasons for the shoplifting which should be addressed beyond any judicial punishment for the crime following apprehension.

so happens that the only group left is teenagers. While this may seem like cynical word-play (after all the effect is the same) it demonstrates the importance of perceptions in understanding stigmatization. That is, there may genuinely be no intention to stigmatize, but the result might be such that those effected by the decision feel stigmatized nonetheless. Furthermore the consequences of the perceived stigmatization might be that more teenagers begin to shoplift (thinking that they will be treated as shoplifters either way). As such the unjustified stigmatization might serve to encourage shoplifting rather than reduce it. However this is speculative and ultimately an empirical question. I mention it here only as a possible outcome.

As with scenario (ii) there will, in scenario (iii), be a question of how many of the innocent will feel stigmatized and harassed. Imagine Diebesstadt I and II, cities of identical population and age distribution to Cleptopolis I and II. In Diebesstadt I there are 1,100 shoplifters, all of whom are teenagers. By focussing surveillance on teenagers there are then 900 innocent teenagers who are treated as suspicious and harassed unjustifiably. In Diebesstadt II, however, of the same 2,000 teenagers 1,999 of them are shoplifters. In this case there is just one teenager who is innocent. Once more the numbers make Diebesstadt II very close to scenario (i) as (almost) every and only teenagers shoplift. The consideration of surveillance of teenagers in Diebesstadt II then turns on the importance placed on the one innocent's right not to be stigmatized and harassed.

Finally there is scenario (iv) in which not every teenager and not only teenagers steal from shops. Of the four this is the only case which exists in the real world. If targeting teenagers when not only teenagers shoplift leads to ineffective surveillance, and targeting teenagers when not all teenagers shoplift both stigmatizes and harasses the innocent, then in scenario (iv) such targeting risks both ineffective surveillance and the stigmatization and harassment of the innocent. As with scenarios (ii) and (iii), the numbers here also matter. The closer the numbers approach scenario (i) of every and only teenagers shoplifting the more effective will be the surveillance, the more justifiable will be the stigmatization and the fewer innocents harassed. By contrast as the numbers involved fall to levels of many, some or just a few teenagers shoplifting, so these concerns become more problematic and less justifiable. As above, much of the justification will turn on the importance placed on the individual's rights not to be harassed or stigmatized, especially when weighed against society's desire to be free from shoplifting.

A final problem is that more realistic versions of scenarios (i-iv) are unlikely to remain static over time. Hence even if scenario (i) were to exist in a particular city for a particular year, the following year the statistics may change such that scenario (iii) now prevails. Hence in one year a policy such as that suggested above may be devised based on the justified stigmatization of teenagers, owing to the fact that they are all known to shoplift. In a subsequent year, in which the individuals making up the population of teenagers has changed such that not all teenagers now shoplift, the policy would be based on the *un*justified stigmatization of teenagers. The change in the justification of the policy would have nothing to do with the policy *per se*, but rather changes in the group it is singling out for attention.

In summary I have argued that there are four alternative scenarios with reference to teenagers and shoplifting:

i) every teenager and only teenagers engage in shoplifting;
ii) every teenager but not only teenagers engage in shoplifting;
iii) not every teenager but only teenagers engage in shoplifting; and
iv) not every teenager and not only teenagers engage in shoplifting.

Of these, scenarios (i) and (ii) were seen to be unproblematic in terms of unjustified stigmatization and harassment of the innocent. That is, in cases where every teenager engages in shoplifting the profiling of such teenagers does not lead to these harms. Depending on the numbers of non-teenage shoplifters in scenario (ii), though, the profiling may prove to be inefficient in that it could fail to identify the majority of shoplifters. Scenarios (iii) and (iv) were seen to demonstrate unjustified stigmatization and harassment of the innocent. It was notable that this could be the case, at least in the perception of some, even where no stigmatization was intended. Finally I also commented that changes may occur between the scenarios over time. Such changes could imply that a policy which was at one time justified may become unjustified, and would need to be altered accordingly.

1.2.1.2 Group Profiling in Reality

The above cases of Cleptopolis and Diebesstadt are unrealistic and overly simplistic, yet they clarify major problems with profiling based on group membership. In more realistic cases when the numbers are far less extreme these problems will be exacerbated. Furthermore, surveillance is rarely if ever 100% effective and it might not be warranted in particular cases. One might think here of well-publicised cases of the surveillance of dog walkers to apprehend those who do not clean up after their animal has defecated.

In addition to these concerns, there is a further issue of self-fulfilling prophecies. Through watching teenagers more closely, more shoplifting teenagers will be caught stealing and sentenced. As more teenagers are sentenced for shoplifting, so the statistics will show that a disproportionate number of shoplifters are teenagers. This will then justify the further concentration of attention on teenagers and so on. Meanwhile, other (non-teenage) shoplifters will continue to go undetected by the surveillance which increasingly ignores them. Those who stigmatize teenagers will feel affirmed in their prejudice and continue to indulge it.

Identification of threat based on group identity is therefore likely to be problematic in most real-life cases. It introduces or perpetuates social stigma, places a burden of suspicion on the innocent and risks instituting self-fulfilling prophecies. Furthermore such group identification is rarely limited to particular age groups, as in the above examples. It often includes ethnicity and religious identification, especially when related to crime and particularly terrorism (Warikoo 2011). As a result it seems as if there should be a presumption against group profiling unless it can be demonstrated that a) a significant majority of the group deserve to be subjected to surveillance (although quite how many form a "significant majority" is unclear); b) the ensuing wrongs outlined above are worth paying when balanced against the benefits to be had from the surveillance; and c) the surveillance is both effective and warranted in the particular case.

1.2.2. Behavioural Profiling

What then of identifying threats according to behaviour? Were the target group to act in a distinctive way then they could again be singled out for attention while the majority are unaffected. However such distinctive behaviours seem hard to find. When operators and the police have restricted their attention to suspicious behaviour this notion has often been ill-defined and of limited value (Police and Criminal Evidence Act 1984; Graham 1998). Related problems have been noted in police "stop and search" tactics which require "reasonable grounds for suspicion". In practice these evince a high degree of prejudice against particular age groups and ethnicities (Dodd 2010). This may be because these groups display behaviour which is misinterpreted as suspicious by police, particularly when the police are predominantly drawn from a different age or ethnicity. However it may also be due to the poor definition of "suspicious" meaning that behavioural profiling can serve as a mask for continued group profiling. As such the problems of prejudice once more come to the fore.

Even when it is not masking group prejudice *per se*, behavioural profiling can still involve such prejudice. As noted above, different age and socio-ethnic groups are known to display different behavioural characteristics. One only has to think of the difficulty many adults have in communicating with teenagers to realise that what is normal behaviour in one age group can be abnormal in another. By defining a particular behaviour or characteristic as identifying a threat, there is a risk that one also inadvertently identifies an innocent group which uses that behaviour in a non-threatening manner. Consider here large groups of teenagers hanging around outside McDonalds with their hoods up and their heads bent. For many adults this carries a lurking sense of threat, even when no threat is intended on the part of the teenagers. Even more directly, the behaviour could be that which identifies a group. Hence while some teenagers wear their hoods up and others do not, it might be that all members of a particular group perform a particular action (say, going to a religious building on a particular day). As such the behaviour might be that which is explicitly profiled, but the effect would be the same as in group profiling.

1.3 False Positives and False Negatives

Limited processing capacity and social stereotyping together contribute to the problem of false positives and false negatives. False positives in this context are people who capture the attention of the operator even though they have done nothing wrong. An example here would be an alarm placed at the exit to a shop in order to detect those leaving the shop without paying for an item. Each person who sets off the alarm, even though they had paid for all of the items they were carrying, would be a false positive. The system would incorrectly identify that person as a shoplifter. By contrast if a person really was carrying an item for which they hadn't paid and walked passed the alarm without setting it off then that person would be a false negative. That is, the system would incorrectly register that person as a non-shoplifter (or, more correctly, it would fail to register that person as a shoplifter).

There is some ambiguity as to when something or someone becomes a false positive or a false negative. The target needs to be clearly defined to all concerned. For example, an airport security system might target terrorists, but this needs further clarification. Leaving aside the definition of "terrorist", is the target:

1) terrorists about to destroy an aeroplane (TD), or
2) just any terrorist who happens to be passing through the airport with no intention of destroying an aeroplane on this visit (AT)?

If TD is the target then AT would be a genuine negative for that system. It is not designed to uncover AT and so should not be judged on those grounds. While the target is simply defined as "terrorists" however, without clarifying whether "terrorist" refers to TD or AT, this definition is ambiguous and open to misinterpretation. As such the definition of a false negative is relative to the aims and purposes of the system.

Similar problems affect the definition of false positives, which again rely on clear and unambiguous definitions of the target. If the security system is designed to locate AT then any terrorist will be a target. If, on the other hand, it is designed to locate TD then any terrorist flying to see his mother with no intention of an attack on this occasion would be a false positive, albeit a welcome find.

There is a further problem affecting false positives in that there are often stages of filtering before a final decision is made. Keeping with the airport security system we might say that a false positive is any innocent who is incorrectly arrested as a terrorist. I will call this a final false positive. If, though, the stages of filtering are taken into consideration then there will be false positives at each stage. Indeed, the system might be designed specifically so that there are false positives, hence:

- Stage 1 – CCTV operators look for suspicious individuals;
- Stage 2 – CCTV operators take a sustained look at individuals from Stage 1;
- Stage 3 – ground-based agents take a sustained look individuals passing Stage 2;
- Stage 4 – ground-based agents remove individuals passing Stage 3 for interview;
- Stage 5 – the arrest or release of any individuals from Stage 4.

Stage 5 is hence the final stage from the perspective of that security system. In this case the successive filtering is designed to accommodate false positives, albeit fewer at each stage. Hence Stage 1 might involve 1,000 false positives, Stage 2 involve 500 false positives and so on until Stage 5 has relatively few false positives. Each of the stages prior to the final stage therefore has its own false positives, or non-final false positives. Furthermore, while the arrest of an innocent is obviously regrettable, this may be felt to be a price worth paying for the evil avoided, provided those innocents are then recognized as such following their arrest. Thus Stage 5 is the final stage in the surveillance, but it is not the final stage from the perspective of the broader aim of capturing and imprisoning the guilty.

As a system might be designed to accommodate false positives at different stages to allow for progressive filtering it is not the case that false positives are inherently problematic. They may even prove to be beneficial, as I shall argue in Section 3.1. The area in which problems arise is when there is a cost. Take first the cost to the subject under surveillance who turns out to be a false positive. In the above case Stages 1-3 impose comparatively little cost on the individual under surveillance. That he is identified as a potential target at Stage 1 and then dismissed as a false positive at Stage 3 might occur without his knowledge. This is not to say that such surveillance is cost free. The suspect may suffer from a violation of privacy or an unknown harm as a result of the surveillance. However the cost incurred increases

significantly at Stage 4, the interview, when the suspect is inconvenienced and is likely to feel harassed. At this stage it is more costly to be incorrectly identified as a genuine target and hence the presence of false positives in the system becomes more problematic.

The question of costs cuts both ways. While there are costs to the surveilled subject there are also costs to the surveillant. Operationally, the greater the number of false positives, the greater will be the difficulty in finding the genuine cases. This is akin to finding the proverbial needle in a haystack, the false positives contributing to the amount of hay. The number of *final* false positives may be mitigated by increasing the stages of filtering, but this is costly in terms of time and resources. Furthermore the later, more interactive stages (i.e. those involving ground-based agents) are likely to be more resource-intensive. As such it would be preferable for the surveillant to reduce the number of non-final false positives before this stage, especially if that can be done with minimal cost to both surveilled subject and surveillant.

Moving to false negatives, it is possible to say that, to some extent, every false negative is a failure of a particular system. Each is an example of one who "got away". However no system is perfect and so some false negatives are to be expected. Nonetheless we should aim to reduce the degree of error. The degree to which we should do this will once more depend on the cost of the harassment and stigmatization of the false negative when weighed against the cost of the wrongdoing which the surveillance is intended to address. If for instance the intended targets are terrorists about to blow up a plane then one false negative is likely to be one too many. If, on the other hand, the target is a person entering a country without valid travel documentation because he is seeking work as an illegal alien then this, as an isolated incident, is less problematic.

Given that there may be several security systems in place it is possible to distinguish between final and non-final false negatives in the same way as with false positives. If one is considering a single system (taken out of the context of the overall complex of systems) then every false negative is likely to be a final false negative. Once a person has been eliminated from the system it is unlikely to spend any more time processing him as a potential threat. If considering a system in the context of an overall complex of systems, however, a non-final false negative is less problematic if the target is located by a different system. He would only class as a final false negative if he passed through *every* system undetected.

When linked with group membership, false positives can contribute, as I have shown, to harassment and stigmatization. A significant number of false negatives imply that the system is not working, and so there are relatively few benefits against which to balance costs. Ideally, then, a system will seek to reduce final false positives and final false negatives as much as possible. To do this, though, it may seek to increase non-final false positives and non-final false negatives provided they come at a relatively low cost.

1.3.1 SPOT

A helpful illustration here is the US Transport Security Administration's (TSA) Screening Passengers by Observation Techniques (SPOT) programme. SPOT operates at certain airports and involves the behavioural profiling of passengers, looking for "facial expressions, body language, and appearance that indicate the possibility that an individual is engaged in some form of deception and fears discovery" (Lord 2010). Deploying 3,000 officers to 161 US airports, the TSA is estimated to have observed 2 billion people between May 2004 and August 2008. Of these, 151,943 were subject to secondary screening, 14,104 were then interviewed, and 1,083 were arrested (Lord 2010; see also Mica 2010). During this period the US Government Audit Office (GAO) believes that "at least 16 of the individuals allegedly involved in [terrorist] plots moved through 8 different airports where the SPOT program had been implemented … on at least 23 different occasions" (Lord 2010). None of the sixteen were apprehended by officers involved with SPOT. Following the GAO report there was discussion in the media and by politicians over the "failure" of SPOT to apprehend a single terrorist (cf. Harwood 2010; Keteyian 2010). This, however, is a flawed response for a number of reasons.

The definition of false positives/negative depends as I have shown on the purpose of the system. The stated purpose of SPOT is to help identify "persons who may pose a potential security risk at TSA-regulated

airports by focusing on behaviours and appearances that deviate from an established baseline, and that may be indicative of stress, fear, or deception" (Lord 2010). In particular, SPOT was intended to "deter terrorists" and "counter terrorist activities" (Lord 2010). It did not limit itself to terrorism *per se*, though, and included criminals posing a risk as a target (Lord 2010). Ignoring the ambiguity in the term terrorist, a "potential security risk" could thus fall into one of three categories: a terrorist about to destroy a plane (TD), any and all terrorists (AT), and non-terrorist criminals (NTC).

Given that it was a surveillance programme based on recognizing suspicious behaviour and leading to interviews, the final stage of SPOT should be considered those referred to interview. Final false positives would then be those referred to interview but not subsequently arrested. If SPOT had been intended to catch *just* TD or AT it therefore produced 14,104 final false positives, namely those referred to interview but not arrested for TD- or AT-related offences. Given that it was intended to catch NTC as well, though, these arrests should count as genuine positives.[2] If so then the number of final false positives was 13,021, that being the number of people identified for interview and not arrested. This is noteworthy for it is at this stage that a significant cost was levied on both surveilled subject and surveillant in requiring an intervention and interview to take place. In terms of non-final false positives, i.e. those identified for secondary screening but not subsequently arrested, the number was 150,860.[3]

The SPOT programme therefore produced a large number of (non-final and final) false positives. These were processed at a cost to both the TSA and those selected for interview, for comparatively little gain (1,083 arrests, none of which was for a terrorist-related offence). At the same time it is known to have missed 16 people who would be classed as AT. SPOT therefore does not recognize AT but it is successful in recognizing at least some NTC. However uncovering AT would be ambitious to the point of fantasy. Through profiling behaviour there seems little reason why SPOT should locate any terrorist using the airport for innocent purposes (i.e. to visit their mothers). The alleged terrorists may not have evinced any suspicious behaviour and so be undetectable by this method. It is, I have argued, fallacious to judge SPOT against catching AT unless this was its purpose. However the stated purpose was ambiguous and open to misinterpretation. Apprehending AT might therefore have been the purpose or it might not. If it was, then it was an unrealistic purpose. If not, then SPOT should have been clearer as to exactly what its purpose was. Either way, the ambiguity in its purpose seems to be at the heart of the criticism. By contrast TD is a more realistic goal, but owing to the rarer circumstances of terrorists blowing up (*vice* travelling on) planes, it is harder to evaluate success against this criteria.

1.3.2 Conclusion

Group and behavioural profiling therefore run the risk of creating false positives, resulting in social stigmatization, harassment, and inconvenience. The number of false negatives may also fuel concerns that the system would rack up these costs for limited or no gain. Finally the number of false positives and false negatives will be determined in part by the purpose of the system. It is therefore important that this purpose be spelled out clearly and precisely. If SPOT had a fault, this was it. The wording of its purpose apparently allowed some to believe that it was a means of capturing any terrorist who flew from a participating airport.

2 Automated Surveillance

In manual surveillance the limited processing capacity of the operator may cause him to filter and prioritize information based on irrelevant and prejudicial social stereotypes. Taken together, limited processing capacity and prejudice can inflate the number of false positives/negatives recognized by the operator, which

[2] Given that 39% were illegal aliens, 19% had outstanding warrants and 16% were in possession of fraudulent or suspect documents, however, it is questionable as to how much of a threat these individuals posed to airports.

[3] An alternative calculation here would determine the false positives at this stage being those identified for secondary screening but not referred to interview (i.e. 138,922) *vice* those identified for secondary screening but not arrested.

has implications for the efficacy of the system. Were this the only approach to surveillance these limitations might be considered an acceptable cost. However alternatives have been created through the integration of artificial intelligence with surveillance systems. There is now the possibility of a fully-automated surveillance system in which a computer filters and processes information, and takes action accordingly.

Between the extremes of manual and fully-automated systems there is also the alternative of partial automation. I will therefore work with a scale which moves from manual to fully-automated threat assessment such that the following definitions apply:

- manual (operator filters information and operator decides),
- partial automation (computer filters and operator decides), and
- full automation (computer filters and computer decides).[4]

I now compare a fully automated system with manual surveillance. Having done this I shall turn to consider partial automation as lying between the two ends of the scale presented by manual and fully automated surveillance.

2.1 Processing Capacity

In light of the above discussion there are a number of benefits to full automation. In removing the operator one simultaneously removes operator error. However, to conclude that a fully automated system is free from limitations in processing capacity would be too fast. Despite the incredible speed of processing in computers, this is still less than the human brain. Furthermore the computer can only process the information its code allows it to recognize. For example, a computer might be programmed to recognize a person bending near a car for a period of time. Its code might then cause it to sound an alarm on suspicion that the person is attempting to break into the car. However the computer may not have the capacity to recognize whether that person is in fact bending to have a coughing fit, stroke a passing dog or tie a shoe lace. The common sense that humans generally take for granted is indicative of greater processing capacity than any computer currently has, or will likely have for the foreseeable future.

2.2 Prejudice

Emrys Westacott has claimed that machines can be fairer than people. "Highway police issuing speeding tickets, being human, are unlikely to be completely consistent and impartial. Their decisions may be affected by the race, sex, class, age, appearance, and manner of the people they pull over. Machines that clock speeds, identify license plates, and issue tickets accordingly will be unaffected by such things" (Westacott 2003). If Westacott's analysis extends across all machines then this presents a major advance over the human operator in terms of prejudicial behaviour. However the situation is not as clear-cut as might at first seem.

At the heart of any computerized system is the code which defines how that system will function. That code is written by an author, though, and the author may be as prejudiced as any operator. As the code replaces the operator there is hence a danger that the operator's prejudices will be replaced by those of the author. Several commentators have noted that the values of the author, wittingly or not, are frozen into the code, effectively institutionalizing those values (Agre 1994; Bowker 2000; Lyon 2002). Further, while the prejudices of the operator affect a limited number of surveillance subjects (those coming under the view of that operator), those embedded in the code affect every person subject to the system. This risks institutionalising the prejudice.

No author of code exists in social isolation. Rather he must rely on research, data and models designed by others, and earlier code which may also contain bias. In describing the behaviour of shoplifters in

[4] Other scales have been proposed, notably that described by Endsley & Kiris (1995). However the Endsley-Kiris scale focuses on aspects of decision-making, while my desire is to focus on the filtering of information.

Cleptopolis I, a code's author may thus inadvertently codify the behaviour of Cleptopolis I's teenagers, simply because it is teenagers who are most often filmed shoplifting and so teenagers form a substantial basis of the data set from which the author must work. This data set might simply reflect the prejudices of the camera operators, who for personal reasons related to being ostracized at a crucial time in their lives tend to focus their attention on teenagers. As seen, this could lead to the successful apprehension of 40% of shoplifters but will fail to capture 3,000 shoplifters. Worse still, it could fail to recognize this failure and appear to be a success. The code and the operator's prejudice might thus work together to become a self-fulfilling prophecy regarding the untrustworthiness of teenagers.

Behavioural profiling in manual assessment was shown to be limited by underlying prejudice, misunderstandings and difficulties in defining the parameters of suspicious behaviour. However a computer needs more parameters, and more precise parameters, than a person. This opens the way for more accurate, if more basic, methods of behavioural profiling. To illustrate this, and the attendant ethical concerns, I will draw upon recent work by Hogg and Sochman in using the behaviour of individuals in crowded scenarios to determine whether they are part of a social group (Sochman & Hogg 2010). The purpose of this was to enable a surveillance system to distinguish between a bag which is left in a public place with an associate of the bag's owner and one which is genuinely unattended. Hence genuinely unattended baggage will be recognized as such and set of an alert, while bags left with an associate will be ignored.[5]

The model used by Hogg and Sochman to understand the behaviour of individuals is a modified version of the Social Force Model of Helbing and Molnar (Sochman & Hogg 2010; Helbing & Molnar 1995). In this a combination of four factors are used to generate a prediction regarding associations: the attractive force provided by the goal of the individual, the attractive force keeping those in the same group together, and the repulsive forces of stationary objects and of individuals not in the same group (Sochman & Hogg 2010).

There is limited research into cultural differences in crowds, but what there is indicates significant differences in how people from different cultures behave.[6] Different walking speeds and attitudes towards personal space have been noted and attributed to cultural distinctives (Wiseman 2006; Helbing et al. 2007; Xiaoping et al. 2009). The mean walking speed of 65m/min in Riyadh is considerably less than that in western states such as Canada, where the mean speed is 84m/min (Koushki 1988). Similarly, significant differences have been found to exist between Germans and Indians in terms of tolerated personal space in high-density crowd scenarios (Chattaraj et al. 2009). Both relative speed and considerations of personal space are significant as they have an impact on the repulsive forces used by Hogg and Sochman (Steffen & Syfried 2008). A related phenomenon is lane formation in crowded situations, in which cultural differences have been observed regarding a preferred side, the kind of lane and the order observed (Schadschneider et al. 2008). Once more, this aspect of lane formation as cultural phenomenon is significant in determining repulsive forces used in the code.

Alongside culture there is also a difference between the mean walking speeds of men and women. This is explicable in part by cultural factors such as dress and laws affecting access to transport, but also by physique: men tend to be taller and have a longer stride than women (Tanaboriboon et al. 1986). A third area of difference is age. Tanaboriboon et al. demonstrated that while secondary school children in Singapore had a mean walking rate marginally faster than the mean adult pedestrian speed, the elderly had a considerably slower pace than either adults or children (Tanaboriboon et al. 1986; Morrall et al. 1991).

---

[5] It is important to note that the work of Hogg and Sochman was developed for use in a partially automated system. Its inclusion here is purely illustrative of the possibility of unintended prejudice entering a fully automated system and is hence not intended to be critical of Hogg and Sochman.

[6] "To our knowledge, however, the characteristics of the motion of pedestrian groups have not been empirically studied so far. It is basically unknown how moving group members interact with each other, with other pedestrians and with other groups. It also needs to be studied how such groups organize in space and how these spatial patterns affect the crowd dynamics. This is expected to be important for the planning of pedestrian facilities, mass events and evacuation concepts" (Moussaid et al. 2010).

The implications are that people from different cultures, sexes and ages will behave differently in crowds. While most of these differences are seen in walking speeds, evidence suggests that such difference are accompanied by distinct approaches to personal space and tolerance of dense crowds (Chattaraj et al. 2009).[7] It is therefore likely that similar differences in approach to personal space might be found not only across cultures but also across sexes and ages. While the Social Force Model accounts for human interactions between stationary objects and between each other, it is imperative that the data also account for cultural, age and engendered differences. The range suggested for the repulsive forces of walls and other individuals may thus be suitable for Western middle-aged men, yet be inappropriate for women, the elderly, or non-Westerners. Were this range to be built into the surveillance system, these people could register as deviating from the norm owing to stronger or weaker repulsive forces than expected.

The system described risks recognizing social associations between Westerners but not between non-Westerners. This could lead to a disproportionate number of non-Westerners being identified as security risks. The prejudice of the operator is thus here replaced by the unwitting prejudice of a particular code, and in so doing will impact all who are surveilled by this system. To avoid this one would need to ensure that the initial modelling includes consideration of groups from a wide variety of cultures, ages and both sexes. This would be of particular importance in areas of international transit such as airports where there is likely to be just such a variety of people using the area.

The picture of automated systems painted by Westacott as pure in terms of prejudice is hence attractive but limited in its application, as Westacott himself seems to acknowledge. While it might apply to some it does not apply to all automated systems. The above scenario demonstrates how prejudice can effectively continue within such a system.

There are, however, positive considerations which should be taken into account. Firstly, prejudice which is genuinely unintentional is arguably less insulting than that which is intended. This does not deny that effort should be made to eradicate such bias before a system is made operational (just as effort should be made not to impose such a system on someone of a different age, sex or ethnicity in society by accident). However when such insults occur out of ignorance they are surely less offensive than when based on prejudicial opinion. Furthermore in the event of a genuine coincidence there may be no insult at all. If a system could be created which recognized terrorists by their gait, and it just happened that teenagers not involved in terrorism and terrorists (and no-one else) shared this gait then it would be no *insult* to teenagers not involved in terrorism that they were inadvertently targeted by the same system which targets terrorists.

Secondly it might be easier to correct a prejudice in the code than in an operator (or numerous operators in the case of institutionalised prejudice). If so then some recoding can affect the entire system in a simple roll-out procedure, eradicating the prejudice overnight. A more challenging scenario would be the case when prejudice in the code cannot be resolved. This returns us to the position considered above in which the benefits of surveillance had to be weighed against the costs of stigmatization, harassment and self-fulfilling prophecies.

2.3 False Positives and False Negatives

As with manual surveillance, the limited processing capacity of a computer and the possibility of prejudice within the system influence the number of false positives/negatives. The number of false positives could increase in an automated system owing to the reduced number of parameters with which the computer can deal. I have argued that whereas the human brain is capable of processing a vast array of data in remarkably little time, the only data with which computers can cope is that allowed for in their code. Let us return to the simple code for behavioural profiling: bending near a car for a period of time is suspicious and so triggers an alert. An innocuous event such as someone bending to tie up his shoe next to a car would then be suspicious to an automated system when it may not to a human. To overcome this false positive, the additional action of tying a shoe needs to be entered into the code. However we saw that there

[7] "High density" in this context being defined as more than one person per square metre.

are other reasons to bend down, such as patting a dog or having a coughing fit. Each of these possibilities needs to be recognized by the system. The likelihood of encountering false positives, even when using behavioural rather than group profiling, is therefore high.

The scope for false positives is extended by the potential for prejudice within the system. As with the human operator, filtering based on prejudice which is irrelevant may return false threats. Taking the illustration of behavioural profiling, non-westerners may be more readily recognized as threatening by the system simply because they walk in ways which that system does not recognize as associative. Where a Westerner is recognized as leaving a bag with a friend, the non-Westerner is seen as leaving unattended baggage and so triggers an alert.

False negatives similarly remain an issue in automated systems, and their number may also be increased by automation. It is likely impossible to develop a code which will describe all possible behaviours and correctly interpret the intentions underlying those behaviours. History has also shown a remarkable ability of ill-intentioned people to adapt their behaviour precisely to avoid detection. Furthermore attempts to profile behaviour indicative of guilt such as SPOT are of little help if an unwitting person is used to carry a bomb onto an aircraft. Generally this is recognized and dealt with by avoiding over-reliance on one form of screening. However the short-comings of new methods with novel technological aspects are easily forgotten in the day-to-day, or in the rhetoric of salesmen and politicians.

3 Partial Automation

I have so far considered manual threat assessment and fully automated threat assessment as opposing ends of a scale. In both cases a combination of limited processing capacity coupled with the potential for filtering based on irrelevant prejudicial criteria increased the capacity for false positives and false negatives. I turn now to consider partial automation. Here I will assume that the purpose of the system is to recognize suspicious behaviour and alert an operator to this. This is not necessarily the case as extremely expensive equipment might render it more cost-effective to have humans perform the front-end filtering. However this is less common than the situation on which I shall focus. The defining feature of any partial automation is that a human element is retained at a real time juncture in the process, rather than after the event as an auditor.

There are two common alternatives in partial automation concerning how the information is presented to the operator. One is for the computer to search for particular suspicious behaviours while leaving the operator free to view the same number of screens as with manual surveillance, with reduced attention paid to the behaviours sought by the computer. I shall call this option “operator as unblinkered” or simply “Unblinkered”. The alternative is for the computer to restrict the operator’s vision to just those screens displaying behaviours it finds suspicious, using the operator as a second filter. This I shall call “operator as blinkered” or “Blinkered”. The key distinction is the scope of the operator to find information not recognized by the computer. In Unblinkered the operator has the same visual scope as in manual surveillance, whereas in Blinkered his vision is restricted to what the computer wants him to see.

This section will consider both Unblinkered and Blinkered partial automation as they compare with manual and fully automated surveillance. Through the combination of operator and computer does one end up with the best or the worst of both worlds? I shall argue that both cases of partial automation are usually preferable to the alternatives.

3.1 Processing Capacity

One of the central concerns of manual operation looked at above is the operators’ need to impose arbitrary filters on an excess of information. Partial automation reduces the quantity of information reaching the operator (Blinkered) or prioritizes that information for him (Unblinkered). Partial automation therefore also reduces the need for the operator to impose his own, potentially flawed filters.

A second benefit is a reduction in boredom, inattentional blindness and change blindness. Where actions are defined in the code of the computer, it will not "miss" them due to *ennui*, fatigue or blindness. The overall improvement in terms of attention will not be as great as in full automation, however. There is still some reliance on the operator who may miss or misread alerts given by the computer.

Partial automation provides increased processing capacity over the alternatives by combining the "unblinking eye" of the computer with the increased mental ability of the operator. To answer the best/worst of both worlds question, the strength of the computer lies in its tireless watching and basic recognition functions. These are precisely the areas in which the operator is weakest. By contrast the operator's strengths lie in more advanced detection of intentions which are beyond the scope of current computing technology. Hence provided the computer handles the primary level of filtering, and is actually capable of handling that level of filtering, and the operator functions at the secondary level there ought to be a best of both worlds scenario.

Of the two alternatives within partial automation, Unblinkered duplicates the work of the computer with the operator. He may see events which the computer will also recognize. The strength of Unblinkered given current computing capacity is that the computer, as noted above, is still very basic. It can only recognize a limited number of human behaviours. Unblinkered therefore allows for the operator to supplement the computer in recognizing more suspicious behaviours than the computer alone. With time and development in computer capacity, however, this looks set to change. As computers become more able to recognize basic behaviour, so the information returned to the operator may be reduced. This would allow the operator to focus on his strengths rather than devote energy to duplicating the work of the computer. As things currently stand, therefore, Unblinkered is a preferable option in terms of processing. However development will lead in most cases to Blinkered becoming a more efficient approach in the longer term.

It is worth stressing that the preference for Blinkered over Unblinkered will be on a case by case basis. There are feasible scenarios in which it will not be necessary to develop the more filtered approach envisioned by Blinkered. For example, automated chemical "sniffing" of luggage to detect explosives is currently at a reasonably advanced level. An operator may stand at a distance while equipment analyses passing luggage for explosives, the operator responding to alerts as received. It is possible that the filtering system returns a low number of false positives daily which are easily handled by the operator. While development in processing capacity might further reduce these false positives, such development may be seen as unnecessary.

Partial automation can also introduce complacency, a fundamentally new concern which is absent in the alternatives (Parasuraman et al. 1993). Should the operator believe that the computer functions effectively without him he may pay less attention to the decisions the computer suggests. The unblinkered operator may cease to notice threats not recognized by the computer. The blinkered operator may likewise simply authorise all suggestions without checking them adequately. In both scenarios there is a risk of too much faith being placed in automation by the operator. In this instance allowing a greater number of non-final false positives (in which final refers to the post-human stage) to exist might prove to be beneficial in reducing the operator's temptation to rely on the automation. If it is known that 60% of the cases flagged up for his attention will be (non-final) false positives then it might prevent his becoming too complacent. Hence in some cases the number of non-final false positives might not only be manageable but also preferable.

I shall return to consider false positives and false negatives in Section 3.3. It is important to note at this stage however that the above examples demonstrate the case-by-case basis for preferring one form of partially-automated surveillance over the other. The considerations of context, purpose, complacency and cost will all affect whether the more blinkered approach should be pursued.

3.2 Prejudice

As with full automation, partial automation can reduce prejudice. While some prejudice may remain in the code I have argued that this may be remedied more easily in an automated system than in a human.

Regarding prejudice, the combination of operator and computer might lead to one of three outcomes. The operator may notice persistent prejudice in the automated system and alert others to this. Alternatively the prejudice in the system may converge with the operator's own prejudice and so go unnoticed. Thirdly, the operator may choose to ignore the system's analysis, opting to be informed rather by his own prejudiced beliefs than the computer's software. An example of the third alternative would be the operator in Cleptopolis (where every teenager but not only teenagers shoplift) who is disposed against teenagers and so ignores an automated filter showing a middle-aged man shoplifting because "only teenagers do that sort of thing." This would presumably be coupled with a belief that the system had missed some important exonerating information which would justify the apparent theft.

The scope for operator prejudice is less in Blinkered than Unblinkered because the scope of information presented to the operator is restricted. The operator as unblinkered remains free to look for his own "targets" and so to monitor those he chooses. In Blinkered the operator's prejudice is only brought to bear on a target already branded suspicious by the computer. The operator's prejudice can affect the outcome, but it is prevented from wholly determining the individuals who should be considered for that outcome. Overall there is a reduction in scope for prejudicial decision-making through the increased computer recognition of salient factors and the reduced need to arbitrarily filter information.

Crucially the retention of the human operator might therefore reduce or contribute to prejudice. It was noted at the beginning of this paper that the prejudice of operators was highlighted in the research of Norris and Armstrong. To what extent, though, is this research a reliable guide to operator behaviour?

Norris and Armstrong carried out their research in three sites in the UK between May 1995 and April 1996 (Norris & Armstrong 1999). The twenty-five operators at these three sites were monitored for 592 hours, or 74 eight-hour shifts (Norris & Armstrong 1999). The retrieved data concerned "888 targeted surveillances. In 711 of these surveillances, a person was identified and [there is] basic demographic data for each of them (age, race, sex and appearance)" (Norris & Armstrong 1999). During this time, the surveillance resulted in 45 police deployments. Of these, "two resulted in no suspect/target being identified by police on the ground. Of the remaining 43 incidents, 76 per cent resulted in … a warning and those identified were allowed to go on their way. An arrest was made in … 12 incidents, that is in less that one-quarter (24 per cent) of all deployments and less than one in seventy targeted surveillances" (Norris & Armstrong 1999). As Norris and Armstrong state, this amounted to an average of "twelve targeted observations per shift, roughly one every forty-five minutes and deployment … resulted from about 5 per cent of target surveillances" (Norris & Armstrong 1999).

From these initial statistics several aspects become clear. Firstly the research is dated, occurring 15 years ago at the time of writing this paper. This is not the fault of the authors, who published their findings in 1999, but it is notable that no thorough survey of operator behaviour has been carried out since. In the meantime there has been growth of CCTV usage in the UK, partly as a result of government funding in excess of £208m for over a thousand schemes from 1994-2003 (Gerrard et al. 2007). Furthermore, both this money and the Norris and Armstrong research targeted cameras operated by local authorities, which are "a very small proportion of the nation's CCTV provision, since the vast majority are commercially owned" (Gerrard et al. 2007). Hence not only was the research carried out 15 years ago, but it concerned a comparatively small (if significant) area of CCTV surveillance. In addition, this area had only just started to undergo a rapid expansion which would continue for another seven years after the end of the survey. The concentration on 25 operators in just three centres out of hundreds also means that the data risks failing to accurately represent the behaviour of all operators. Owing to these concerns it is questionable how useful the Norris and Armstrong data remains, aside from being the only data available.

This is not to exonerate the operators in the Norris and Armstrong study from blame. There were instances of racist language and a disproportionate targeting of the young, the black and the male. However many of the problems were reducible, as Norris and Armstrong point out, to the relationship between the operators and the local police (Norris & Armstrong 1999; see also Norris 2002). Operator behaviour in their study is therefore as much, if not more, a factor of this relationship than of operator prejudice *per se*.

There is hence a danger in misreading or giving too much attention to the findings of the Norris and Armstrong data. This is frustrating as it is some of the most comprehensive data on the subject available, and yet is nonetheless limited by both time and scope. In the intervening 15 years, the growth of CCTV usage suggests that the situation today may be very different from that encountered in the mid-1990s. The value of this data in critiquing current operator behaviour is therefore limited. It is possible that some operators continue to be prejudiced and erratic when they alert the police to incidents. It is also possible, however, that with growth have come professionalism and experience, leading to a more responsible and reliable operator.

3.3 False Positives and False Negatives

Partial automation reduces the information flow to the operator, either literally or by prioritizing that information, and so limits the need for arbitrary or irrelevant filters. This, coupled with the potential for reduced prejudice, means that partial automation can reduce the number of false positives and false negatives.

False negatives may be reduced in both Blinkered and Unblinkered. In both cases, what the operator alone would have missed, the computer may catch. Similarly the reduced need for filters based on social stereotypes should result in fewer false positives at all stages of filtering. I have argued that there is nothing inherently wrong with false positives so long as they do not impose a cost. Those based on social stereotypes typically do carry such a cost in terms of stigmatization, harassment of the innocent and self-fulfilling prophecies. Those based on behaviour may also carry this cost, but this may be easier to resolve. I have also argued that in responding to context, purpose, complacency and cost, the toleration of some false positives might be beneficial to the system. This is especially true if those are low-cost non-final false positives. It would be preferable to have more rather than fewer low-cost non-final false positives if this proved necessary to avoid false negatives.

To illustrate this imagine a bizarre discovery that healthy suicidal terrorists about to destroy a plane almost always walk at 60m/min, while the majority of non-terrorists walk faster than this. Software could then be developed which targeted people walking at this pace. The operator could function, in theory at least, as a second filter to remove the false positives (non-terrorists walking at that pace) before ground-based staff intervene.[8] However there may be terrorists who limp owing to some prior carelessness in placing bombs and so walk at a slower pace. Rather than miss these limping terrorists it may be worth expanding the range of the software to recognize those walking at 60m/min and slower. This would expand the number of non-final false positives, but the return (recognizing limping terrorists) might mean that the burden of the extra non-final false positives could be felt worthwhile to avoid false negatives.

If this is true of non-final false positives, what of final false positives where the operator's decision is the final stage of the process. Once more this would depend on cost. For example, if the operator's authorisation led to a remote-controlled gun shooting the suspected terrorist there would be a much higher cost to false positives than if his authorisation informs a ground-based agent to interview the subject. In the latter case the stage is final from the perspective of the surveillance, although not from that of the overall process. In the former it is final from the perspective of both the surveillance and the overall process. In either case the impact of a final false positive is likely to be greater than that of a non-final false positive. As such, and while the core issue is still one of cost, it will likely be the case that final false positives should be reduced where possible.

Compared with the alternatives of manual and full automation in terms of efficacy, partial automation is thus preferable if flawed. It has greater processing capacity than either of the alternatives. While the effects of this might be outweighed by prejudice in either the computer or the operator, in neither case is this prejudice irresolvable. As with prejudice in automation, discrimination in software could be recognized and corrected across the system. And as with prejudice in manual surveillance, operator prejudice is local and could be addressed through training and supervision. The more this prejudice, such as

[8] Quite how the operator would do this in practice need not be of concern for the point at hand.

it is, can be reduced the more effective partial automation will become. Thirdly the problem of complacency also needs to be recognized and addressed. However this could likewise be reduced through training, supervision and in some cases the toleration of non-final false positives (to combat operator complacency).

Overall partial automation is therefore the most effective means of surveillance considered. Flaws remain in the system, but when recognized these could possibly be addressed through a combination of training and oversight, as well as allowing for some non-final false positives in the system to counter complacency and to avoid false negatives. Even if these flaws cannot be resolved, though, the partially automated system remains preferable to the alternatives.

4 Further Considerations

Partial automation is the most effective form of surveillance considered. However ethical concerns are not limited to efficacy. What have yet to be considered are privacy and distance, which have bearing on the distinction between Unblinkered and Blinkered. As such, is to these that I now turn to see whether they might provide clarity as to which of the approaches to partial automation is preferable.

4.1 Privacy

Regarding manually-operated CCTV there is often an assumption that the surveillant and the surveilled are anonymous to one another, but this is not always the case. When police monitor known criminals the anonymity is asymmetric. In settings such as CCTV in the High Street the operator, John, might see someone he knows, Jessica, going into a shop selling lingerie and adult toys. While Jessica is entitled to go into that shop, were she aware that John were watching then she might choose not to do so. John is an ex-boyfriend and she would rather he didn't know anything about certain aspects of her current lifestyle. Were John physically present in the street then Jessica would have a reasonable chance of seeing him and avoiding the shop on this occasion. Given John's physical distance, though, Jessica has no means of knowing that he is watching her. An area of her life which she would rather keep private is no longer so, owing to the distance between herself and John introduced by the cameras.[9]

Automated surveillance, by contrast, brings with it a degree of anonymity and privacy. The anonymity was implicitly recognized by Westacott when he noted that machines are less concerned by age, sex, colour, etc. The failure to discriminate between people in any way other than the one under consideration (speeding in Westacott's example) promises a high degree of anonymity. The implications for privacy derive from this in that the automated camera will not follow Jessica's movements in the High Street simply because she is Jessica, in the way it might have done when operated by John. If Jessica's use of the shop is caught incidentally (through random camera sweeps of the area, for example) and the film stored for a period lest needed for evidence, then there is a possibility that someone will see it. However the chances of Jessica's expectation of privacy being violated are reduced in the case of automated surveillance.

Unsurprisingly the threats to privacy in partial surveillance fall between these extremes. The operator as blinkered maintains some access to potentially private information, although this is less than the operator as unblinkered who remains free to access a wider set of information. Indeed, coupled with complacency the unblinkered and immoral operator may feel at greater liberty to focus on privacy-invading activities, expecting the computer to register any important threats. As unblinkered the operator still has free use of the equipment to cast his gaze where he chooses. As such, the privacy risks with Unblinkered are far closer to those associated with manual surveillance. Given that this is a concern for many people with manual surveillance this should be taken seriously. However it is hard to see how to mitigate this. The system is designed to allow the operator free access to the cameras and so the potential for privacy violations at the whim of the operator remains as a part of the system.

[9] Interestingly, in e-mail correspondence, Clive Norris has suggested that pictures of the faces of all relevant camera operators be posted in the areas in which CCTV is operating in an attempt to overcome this problem.

By contrast the blinkered operator has only the information returned owing to the computer recognising suspicious activity. Hence any infringement of privacy would be incidental to such activity. This is virtually identical to the risks associated with auditing fully automated surveillance. This therefore argues strongly in favour of Blinkered, *ceteris paribus*, where it is feasible to respect the privacy expectations of the surveilled subjects.

4.2 Distance

A related concern is the physical distance which exists between the operator and the surveilled subject. Such distance does not occur when, for instance, a policeman confronts a suspect (Norris 2002). This distance takes the imminence out of a situation for the operator, allowing him more time to reflect and seek advice than if he were a ground-based agent. Furthermore, as he is not being threatened he can take a more objective stance than one who is facing a violent person. At the same time distance impacts on the operator's situation awareness, limiting it to what he can see on the screen(s). He might not be privy to aural information which could impact the interpretation of a situation. He may also deliberate for too long over a situation when, had he been on the ground, he would have intervened more rapidly.

A further problem is that distance can grant one's prejudices immunity from being challenged by the realities on the ground. The operator might falsely believe, for example, that an Asian is more likely to commit a crime than a Caucasian. Without face-to-face contact with an Asian person, though, the operator might never have cause to review this opinion. On the other hand, the distance also eliminates some level of discretion which is available to the officer on the ground. This might be an improvement on ground-based enforcement in which stereotyping may play an even greater role than in the control room.

When applied to more automated systems, however, the role of the human surveillant is reduced. Due to the binary nature of computers, one result of automation can be the over-simplification of society. This might not be problematic when determining whether someone has sufficient money on their Oyster card to use the Underground. Were such an approach to become pervasive, however, it could exacerbate the stratification of society between the haves and have-nots. Gaining valid tokens of access (e.g. an Oyster card to use the Underground, which has gone "cashless") might depend on having other tokens (e.g. a credit card), which in turn rely on other tokens (e.g. a place of residence), and so on. While a human system might allow room for subtlety, explanation or pleading, an automated system will not (thus preventing the homeless person from using the Underground). As such the differences in society risk become more cemented rather than more fluid.

As a means of social control, allowing or preventing access to areas of society, Blinkered risks denying the surveilled an opportunity for interaction. This is especially true if his actions are not recognized by the system as suspicious and so are not returned to the operator. This removes the possibility of negotiation, subtlety and discretion from one area of human interaction. While scope for negotiation is already limited in the case of CCTV, the operator does retain some capacity for discretion which a computer does not. A surveilled subject might even make an appeal to a human-operated camera to be allowed access; such appeals would be wasted on more automated systems.

Blinkered therefore risks limiting opportunities for negotiation, subtlety and discretion. To gain the computer's (and thereafter the operator's) attention the surveilled subject might have to act suspiciously. This despite the fact that he has no intention of wrongdoing, and might suffer repercussions for his actions. Without the scope for these interactions one party is disempowered. Without scope for such interaction in an automated system a similar disempowerment may be introduced across the system.

By contrast the expanded scope of the operator in Unblinkered allows for some discretion absent from Blinkered. The consideration of distance can therefore be seen as a companion to considerations of prejudice. As the operator loses discretionary powers he has less scope for making prejudiced decisions but also less ability to make sympathetic decisions. Hence Westacott's example of automated speeding detection referred to above. Discretion thus works both ways. Blinkered, being more automated than

Unblinkered, allows for relatively little interaction and so increases the distance between operator and surveilled subject.

4.3 Privacy vs. Distance

The related considerations of privacy and distance are therefore relevant to partial automation. With greater automation comes greater privacy. However the distance between surveilled subject and operator also increases with automation such that there is less scope for discretionary behaviour or human interaction. Blinkered is preferable to Unblinkered in terms of privacy. Distance, a more morally ambiguous concern than privacy, increases with Blinkered with both positive and negative repercussions. Most notably the scope for discretion is reduced, limiting opportunities for interaction but also for prejudice. There may be a cementing of social mobility, but also less scope for impassioned decision-making.

If concerns relating to distance cancel each other out, but privacy remains an issue, then Blinkered should be seen as generally preferable to Unblinkered. Privacy violations are curtailed and while some negative effects are felt from distance these are counter-balanced by the positive, especially the reduction in scope for prejudice. These are not hard-and-fast conclusions, however. Different cases will preference different solutions. Where human actions are likely to be varied and unpredictable, coupled with poor computing capacity, a broadly homogenous target set and limited scope for privacy violations the unblinkered approach may be preferable. It would be more efficient at recognising threats, require less development, counter complacency and risk less prejudicial decision making based on social stereotypes or voyeurism. By contrast more effective processing capacity of automation coupled with more predictable human behaviour, heterogeneous people groups and/or situations in which a subject might expect greater privacy will lead to situations in which more blinkered options would be preferable.

The limiting factor in Blinkered remains its efficacy based on the code's capacity to recognize suspicious behaviour. As this is currently low it would be better for more information to reach the operator rather than less. Nonetheless, the ethical benefits of Blinkered imply that this, if its processing capacity could be improved, is preferable to Unblinkered. Furthermore, as processing capacity and recognition functions improve, so the quantity of information returned to the operator may be refined and reduced. What remains important, however, is that the human operator is not removed from the system. To do so would be to demonstrate complacency on a societal level in placing too great a faith in the ability of computers to do our work for us. This would, I have argued, result in a less, rather than a more, effective system.

5 Conclusion

In this paper I have considered the ethical concerns arising from automating surveillance. In this I have looked at three alternatives: manual, automated, and partially-automated surveillance. In the case of manual surveillance I argued that there are a number of ethical concerns, not least being the poor processing capacity of the human operator. A tendency to rely on profiling to aid flawed processing, coupled with personal prejudice, was shown to result in likely unjustified stigmatization and harassment, as well as inefficient and potentially costly surveillance.

Manual surveillance was then contrasted with fully-automated surveillance in which a computer filters information and decides on the action to take based on that information. While the computer is free from certain constraints on human operators such as boredom and inattentional blindness, fully-automated surveillance is not free from processing concerns. In particular it was seen that computers may lack a subtlety of awareness which a human would likely describe as common sense. Drawing on research by Sochman and Hogg I also demonstrated that computerized systems were not necessarily free from prejudice. Once more the combination of limited processing and the potential for prejudice indicated that false positives and false negatives would likewise continue to be an issue for fully-automated systems.

I then looked at partial automation in which the computer filters information and passes this to a human operator for action. Two alternatives were considered: blinkered in which the operator only sees that

information presented to him by the computer and unblinkered in which the operator remains free to scan monitors but has his attention alerted to particular events by the automated system. In terms of processing capacity, blinkered was seen to have the advantage of limiting a duplication of effort, but the current state of computer processing led to the conclusion that unblinkered was for the present the preferable option. Both prejudice and false positives/negatives remain a concern for partial automation. In the case of the former I argued that the combination of operator and computer might either reduce or contribute to prejudice. Once more, blinkered was seen to be preferable in this respect, although the limited availability of reliable data meant that the degree to which prejudice was a real issue in surveillance systems is underdetermined. Finally it was noted that false positives and false negatives would be reduced in partially-automated systems.

In conclusion I believe that partially-automated surveillance systems are ethically preferable to either manually-operated or fully-automated systems. They are stronger in terms of processing capacity and, while prejudice and false positives/negatives remain, the concerns seem less significant than those associated with manual surveillance. As to which form of partially-automated system is preferable, it would seem that in a perfect world blinkered would generally be better than unblinkered. However the current state of computer processing means that a blinkered system would likely incur too many false negatives to be effective. Given current conditions, then, it seems as if an unblinkered partially-automated system would, in most cases, be the ethically preferable option.

6 Further Research

This paper has touched on some of the issues arising from the automation of surveillance. Among these I have raised concerns relating to the manner in which prejudice can unwittingly enter an automated system, the potential for which was shown in the case of SUBITO. Designed for partial automation, the SUBITO project developed automated tools for aiding the human operator. In addition to assessing behaviour of individuals in crowded situations, algorithms were also developed to predict behaviour and automate facial recognition. Neither of these are unique to SUBITO (Baker et al. 2008; Huang et al. 2008), and automated facial recognition has been in use for some years (Burrell 1998; Firth 2011). In both cases, though, there is a paucity of published ethical analysis along the lines suggested in this paper.

Secondly, while I have tried to focus on what I perceive to be the central ethical issues of the automation of surveillance, this paper is by no means exhaustive. There are further issues such as the potential for abuse and function creep (Winner 1977). Remaining with the example of SUBITO, could aspects of the technologies discussed in this paper be used to isolate particular ethnic or religious groups? Are there other uses of the SUBITO technology beyond the recognition of unattended baggage, such as locating lost children? If SUBITO were installed to locate terrorists but succeeded only in recognizing lost children and returning them to their parents, would this be a success sufficient to justify the installation of the system? These are questions which I have not had the space to develop in this paper, but which nonetheless require answering.

Finally much of this paper has focussed on the impact that surveillance can have on the individual. I have considered whether the individual may feel unfairly harassed or stigmatized, and how automation might impact upon the privacy of the surveilled subject. What I have not done is to look at the impact of automating surveillance on society at large. Authors such as David Lyon warn of the dangers of social sorting which arise from surveillance (Lyon 2002), while others discuss the chilling effects that can arise in society from high levels of surveillance (Michelman 2009). These harms, if established, must be weighed against the efficacy of the surveillance, and perceptions of that efficacy. While some work has been carried out on perceptions of efficacy (Gill & Spriggs 2005), none has yet addressed how that perception differs in cases of automated CCTV. Finally, it is important to consider the ethical implications of the potential normalisation of surveillance practices. Is it positive for society that we are, or at least presume we are, being watched over by a computer or a person? Similarly, is it ethically advantageous that we depend upon automated surveillance rather than more traditional embodied surveillance in addressing societal concerns such as those outlined above?

References

Agre, P.E. (1994). Surveillance and Capture: Two Models of Privacy. *The Information Society*, 10(2), 101-27.

Baker, C.L., Goodman, N.D. & Tenenbaum, J.B. (2008). Theory-based social goal inference. *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.154.2746. Accessed 11 February 2012.

Bowker, G.C. & Star, S.L. (2000). *Sorting Things Out: Classification and Its Consequences.* Cambridge, Mass.: MIT Press.

Burrell, I. (1998). Face-recognition CCTV launched. *The Independent*. Available at: http://www.independent.co.uk/news/facerecognition-cctv-launched-1178300.html. Accessed 11 February 2012.

Chattaraj, U., Seyfried, A. & Chakroborty, P. (2009). Comparison of Pedestrian Fundamental Diagram Across Cultures. *Advances in Complex Systems*, 12(3), 393-405.

Corby Borough Council Electronic Information Team, The Borough of Corby CCTV Department. http://www.corby.gov.uk/business/towncentremanagement/pages/cctv.aspx. Accessed 17 May 2011.

Dodd, V. (2010). Stop and search plans are "discriminatory", watchdog warns. The Guardian. Available at: http://www.guardian.co.uk/uk/2010/nov/15/stop-and-search-equality-commission. Accessed 31 March 2011.

Endsley, M. & Kiris, E. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381-394.

Firth, N. (2011). Face recognition technology fails to find UK rioters. *New Scientist*, (2826). Available at: http://www.newscientist.com/article/mg21128266.000-face-recognition-technology-fails-to-find-uk-rioters.html. Accessed 11 February 2012.

Gerrard, G., Parkins, G., Cunningham, I., Jones, W., Hill, S., & Douglas, S. (2007). *National CCTV Strategy*, Home Office and Association of Chief Police Officers. http://webarchive.nationalarchives.gov.uk/20100413151441/http://www.crimereduction.homeoffice.gov.uk/cctv/National%20CCTV%20Strategy%20Oct%202007.pdf. Accessed 21 January 2011.

Gill, M. & Spriggs, A. (2005). Assessing the Impact of CCTV. Home Office. Available at: http://rds.homeoffice.gov.uk/rds/pdfs05/hors292.pdf. Accessed 26 July 2010.

Graham, S. (1998). Towards the Fifth Utility? On the Extension and Normalisation of Public CCTV. In C. Norris, J. Moran, & G. Armstrong, (Eds.), *CCTV, Surveillance and Social Control* (pp. 89-112). Aldershot: Ashgate Publishing Limited.

Haggerty, K.D. (2009). Methodology as a Knife Fight: The Process, Politics and Paradox of Evaluating Surveillance. *Critical Criminology*, 17(4), 277-291.

Harwood, M., 2010. Terrorists Slip Past TSA's Scientifically Untested Behavioral Threat Detection Program. *Security Management*. http://www.securitymanagement.com/news/terrorists-slip-past-tsas-scientifically-untested-behavioral-threat-detection-program-007158. Accessed 17 May 2011.

Helbing, D. & Molnar, P. (1995). Social Force Model for Pedestrian Dynamics. *Physical Review E*, 51(5), 4282-4286.

Helbing, D., Johansson, A. & Al-Abideen, H. (2007). Dynamics of Crowd Disasters: An Empirical Study. *Physical Review E*, doi: 10.1103/PhysRevE.75.046109.

Huang, G.B. et al., 2008. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. *Workshop on Faces in "Real-Life" Images: Detection, Alignment, and Recognition*. Marseille, France. Available at: http://hal.archives-ouvertes.fr/inria-00321923/. Accessed 11 February 2012.

Keteyian, A. (2010). TSA's Program to Spot Terrorists a $200M Sham? *CBS Evening News*. http://www.cbsnews.com/stories/2010/05/19/eveningnews/main6500349.shtml. Accessed 17 May 2011.

Koushki, P.A. (1988). Walking Characteristics in Central Riyadh, Saudi Arabia. *Journal of Transportation Engineering*, 114(6), 735-744.

Lippert-Rasmussen, K. (2010). "We are all Different": Statistical Discrimination and the Right to be Treated as an Individual. *The Journal of Ethics*, 15(1-2), 47-59.

Lord, S.M. (2010). *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, Government Accountability Office. http://www.gao.gov/new.items/d10763.pdf. Accessed 18 January 2011.

Lyon, D. (2002). Surveillance as Social Sorting: Computer Codes and Mobile Bodies. In D. Lyon (Ed.), *Surveillance as Social Sorting* (pp. 13-30). Oxford: Routledge.

Mack, A. (2003). Inattentional Blindness: Looking Without Seeing. *Current Directions in Psychological Science*, 12(5), 180-184.

Manning, F. (2011). Worcester City monitor 100 CCTV cameras with only one person. *Big Brother Watch*. http://www.bigbrotherwatch.org.uk/home/2011/08/worcester-city-monitor-100-cctv-cameras-with-only-one-person.html. Accessed 24 August 2011.

McKinnon, R. (2007). Big Brother isn't Watching. *Evening Times*. http://www.eveningtimes.co.uk/big-brother-isn-t-watching-1.976256. Accessed 17 May 2011.

Mica, J.L. (2010). Letter to Janet Napolitano, Secretary, Department of Homeland Security. http://republicans.transportation.house.gov/Media/file/111th/Aviation/2010-05-20-TSA_Reorg_Letter.pdf. Accessed 18 January 2011.

Michelman, S. (2009). Who Can Sue Over Government Surveillance? *UCLA Law Review*, 57, 71-106.

Morrall, J., Ratnayake, L. & Seneviratne, P. (1991). Comparison of CBD Pedestrian Characteristics in Canada and Sri Lanka. *Transportation Research Record*, (1294), 57-61.

Moussaid, M., Perozo, N., Garnier, S., Helbing, D., & Theraulaz, G. (2010). The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE*, doi: 10.1371/journal.pone.0010047.

Norris, C. (2003). From personal to digital: CCTV, the panopticon, and the technological mediation of suspicion and social control. In D. Lyon (Ed.), *Surveillance as Social Sorting* (pp. 249-281). Oxford: Routledge.

Norris, C. & Armstrong, G. (1999). *The Maximum Surveillance Society: The Rise of CCTV*. Oxford: Berg.

Parasuraman, R., Molloy, R. & Singh, I.L. (1993). Performance Consequences of Automation-Induced "Complacency." *International Journal of Aviation Psychology*, 3(1), 1-23.

Police and Criminal Evidence Act 1984.

Resnick, R.A., (2002). Change Detection. *Annual Review of Psychology*, 53, 245-277.

Schadschneider, A., Klingsch, W., Klupfel, H., Kretz, T., Rogsch, C., Syfried, A., & Meyers, B. (2008). Evacuation Dynamics: Empirical Results, Modeling and Applications. In B. Meyers (Ed.), *Encyclopedia of Complexity and System Science* (3142-3176). Berlin: Springer.

Simons, D. & Ambinder, M. (2005). Change Blindness: Theory and Consequences. *Current Directions in Psychological Science*, 14(1), 44-48.

Sochman, J. & Hogg, D., 2010. Who Knows Who - Inverting the Social Force Model for Finding Groups. *IEEE International Workshop on Socially Intelligent Surveillance and Monitoring (SISM 2011).*

Steffen, B. & Syfried, A. (2008) The Repulsive Force in Continuous Space Models of Pedestrian Movement. *Physics and Society,* arXiv:0803.1319v1.

Tanaboriboon, Y., Hwa, S.S. & Chor, C.H. (1986). Pedestrian Characteristics Study in Singapore. *Journal of Transportation Engineering*, 112(3), 229-235.

Warikoo, N. (2011). U.S. ends registration program targeting men from Muslim countries. *The Gazette*. http://www.montrealgazette.com/news/canada-in-afghanistan/ends+registration+program+targeting+from+Muslim+countries/4792096/story.html. Accessed 17 May 2011.

Westacott, E. (2003). Human Oversight of Surveillance Technology. Presentation to the Society for Philosophy and Public Affairs, American Philosophical Association Eastern Division Meeting, Washington DC, 29 December 2003. https://docs.google.com/Doc?docid=0AWI7P4qhQyVvZGY5OW52dmZfMTgyOHpwZHJrZ3Y&hl=en. Accessed 17 May 2011.

Winner, L., 1977. *Autonomous Technology: Technics-out-of-control as a Theme for Political Thought*, The Cambridge, MA: MIT Press.

Wiseman, R. (2006). How Fast is Your City? http://www.richardwiseman.com/quirkology/pace_home.htm. Accessed 20 May 2011.

Xiaoping, Z., Tingkuan, Z. & Mengting, L. (2009). Modeling crowd evacuation of a building based on seven methodological approaches. *Building and Environment*, 44(3), 437-445.