Ethics and Information Technology (2005) 7:75–86
DOI 10.1007/s10676-005-4583-2

*Dis*closive ethics and information technology: *disclosing facial recognition systems*

Lucas D. Introna
Center for the Study of Technology and Organisation, Lancaster University Management School, Lancaster, LA1 4YX, UK
E-mail: l.introna@lancaster.ac.uk

Abstract. This paper is an attempt to present disclosive ethics as a framework for computer and information ethics – in line with the suggestions by Brey, but also in quite a different manner. The potential of such an approach is demonstrated through a disclosive analysis of facial recognition systems. The paper argues that the politics of information technology is a particularly powerful politics since information technology is an opaque technology – i.e. relatively closed to scrutiny. It presents the design of technology as a process of closure in which design and use decisions become black-boxed and progressively enclosed in increasingly complex socio-technical networks. It further argues for a disclosive ethics that aims to disclose the nondisclosure of politics by claiming a place for ethics in every actual operation of power – as manifested in actual design and use decisions and practices. It also proposes that disclosive ethics would aim to trace and disclose the intentional and emerging enclosure of politics from the very minute technical detail through to social practices and complex social-technical networks. The paper then proceeds to do a disclosive analysis of facial recognition systems. This analysis discloses that seemingly trivial biases in recognition rates of FRSs can emerge as very significant political acts when these systems become used in practice.

Key words: biases, disclosive ethics, facial recognition systems, false positives, information technology, politics

Introduction

It would not be controversial to claim that information technology has become ubiquitous, invading all aspects of human existence. Most everyday technologies depend on microprocessors for their ongoing operation. Most organisations have become entirely reliant on their information technology infrastructure. Indeed information technology seems to be a very cost-efficient way to solve many of the problems facing an increasingly complex society. One can almost say it has become a default technology for solving a whole raft of technical and social problems. It have become synonymous with societies view of modernisation and progress. In this paper we will consider facial recognition systems as one example of such a search for solutions.

However, this reliance on information technology also brings with it many new and different kinds of problems. In particular, for our purposes, ethical concerns of a different order. We would argue that information technology is mostly not evident, obvious, transparent or open to inspection by the ordinary everyday person affected by it (Brey 2000). It is rather obscure, subsumed and black-boxed in ways that only makes its 'surface' available for inspection. Imbedded in the software and hardware code of these systems are complex rules of logic and categorisation that may have material consequences for those using it, or for the production of social order more generally (Introna and Nissenbaum 2000; Feenberg 1999; Latour 1992). However, often these remain obscured except for those experts that designed these systems – and sometimes even not to them as we shall see in our analysis of facial recognition systems below. Simply put: they are most often closed boxes unavailable for our individual or collective inspection and scrutiny. This problem of 'closure' is made more acute by the fact that these systems are often treated as neutral tools that simply 'do the job' they were designed to do. Differently put, we do not generally attribute values and choices to tools or artefacts but rather to people. Nevertheless, Winner (1980) and Latour (1991, 1992) has shown convincingly that these tools have inscribed in them value choices that may or may not be very significant to those using them or affected by them – i.e. software programmes are political in as much as the rules of logic and categorisation they depend on reflect or included curtain interests and not others. Enclosed in these 'boxes' may be significant political programmes, unavailable or closed off from our critical and ethical gaze.

Paper prepared for the Technology and Ethics Workshop at Twente

Many authors have realised this and have done a variety of analysis to disclose the particular ways in which these technologies have become enrolled in various political programmes for the production of social order (Callon 1986; Latour 1991, 1992; Law 1991). However, in this paper we would like to ask a different question – the normative or ethical question. How can we approach information technology as an ethical problem? In response to this question we will propose, in accord with Philip Brey (2000), but in a rather different way, that the first principle of an information technology ethics should be *disclosure*. Thus, we want to propose a form of *disclosive ethics* as a framework for information technology ethics. We will aim to show how this may work in doing a disclosive analysis of facial recognition systems. Thus, this paper will have three parts: First, we will discuss the question of the politics of information technology in general; second, we will present our understanding of disclosive ethics and its relation to politics; and finally, we will do a disclosive analysis of facial recognition systems.

The politics of (information) technology as closure

The process of designing technology is as much a process of closing down alternatives as it is a process of the opening up of possibilities. In order for the technology to produce its intended outcome it needs to enforce its 'scripts' on its users. Its designers has to make assumptions about users and the use context and often build these assumptions into the very materiality of their artefacts. These artefacts then function as sub-plots in larger social scripts aimed at 'making society durable' – plots (and sub-plots) that are supposed to generate durable social order in which some ways of being are privileged and others are not. It is this closure that is an implicit part of technology design and use that is of interest to us. Let us consider this closure in more detail.

The micro-politics of the artefact

Technology is political (Winner 1980). By this we mean that technology, by its very design, includes certain interests and excludes others. We are not suggesting that this is always an explicit politics. In fact it is mostly implicit and part of a very mundane process of trying to solve practical problems. For example, the ATM bank machine assumes a particular person in front of it. It assumes a person that is able to see the screen, read it, remember and enter a PIN code, etc. It is not difficult to imagine a whole section of society that does not conform with this assumption. If you are blind, in a wheelchair, have problem remembering, or unable to enter a PIN, because of disability, then your interest in getting access to your account will be excluded by the actual design of the ATM. This 'closure' may not be obvious to the designers of ATMs as they may see their task as trying simply to solve a basic problem of making banking transactions more efficient and accessible. In their minds they often design for the 'average' customer doing average transactions. And they are mostly right – but if they are not, then their biases can become profoundly stubborn. In some senses quite irreversible. Where does the excluded go to appeal when they are faced with a stubborn and mute object such as an ATM? Maybe they can work around it, by going into the branch for example. This may be possible. However, this exclusion becomes all the more significant if banks start to close branches or charge for an over-the-counter transaction (as some banks are doing). Thus, as the micro-politics of the ATM becomes tied to, and multiplied through other exclusionary social practice, trivial injustice soon multiply into what may seem to be a coherent and intentional strategy of exclusion (Introna and Nissenbaum 2000; Agre and Mailloux 1997). Yet there is often nobody there that 'authored' it as such (Foucault 1975; Kafka 1925).

Thus, the politics of technology is more than the politics of this or that artefact. Rather these artefacts function as nodes, or links, in a dynamic socio-technical network kept in place by a multiplicity of artefacts, agreements, alliances, conventions, translations, procedures, threats, and so forth: in short by relationships of power and discipline (Callon 1986). Some are stable, even irreversible; some are dynamic and fragile. Analytically we can isolate and describe these networks (see Law 1991 for examples). However, as we survey the landscape of networks we cannot locate, in any obvious manner, where they begin nor where they end. Indeed we cannot with any degree of certainty separate the purely social from the purely technical means from ends, cause from effect, designer from user, winners from losers, and so on.

In these complex and dynamic socio-technical networks ATMs, doors, locks, keys, cameras, algorithms, etc. function as political 'locations' where values and interests are negotiated and ultimately 'inscribed' into the very materiality of the things themselves – thereby rendering these values and interests more or less permanent (Akrich 1992; Callon 1986; Latour 1991, 1992; Law 1991). Through these inscriptions, which may be more or less successful, those that encounter and use these inscribed artefacts become, wittingly or unwittingly, enrolled into particular programmes, or scripts for action.

Obviously, neither the artefacts nor those that draw upon them simply except these inscriptions and enrolments as inevitable or unavoidable. In the flow of everyday life artefacts often get lost, break down, and need to be maintained. Furthermore, those that draw upon them use them in unintended ways, ignoring or deliberately 'misreading' the script the objects may endeavour to impose. Nevertheless, to the degree that these enrolments are successful, the consequences of such enrolments can result in more or less profound *closures* that ought to be scrutinised. We would claim that the politics of artefacts is much more mundane and much more powerful than most other politics, yet it is often enclosed in such as way to evade our scrutiny. This is particularly true for information technology in which closure is much more powerful as the closure is itself closed off.

On the silent politics of the software algorithm

Having argued that technology is political, we now want to claim that the politics of information technology (in the form of software algorithms) is, in a sense, of a different order (Graham and Wood 2003). We want to contend that scrutinising information technology is particularly problematic since information technology, in particular algorithms, is what we would term an *opaque* technology as opposed to a *transparent* technology (Introna 1998). Obviously we do not see this distinction as a dichotomy but rather as a continuum. As an attempt to draw this distinction some aspects are highlighted in Table 1 below.

Facial recognition algorithms, which we will discuss below, is a particularly good example of a opaque technology. The facial recognition capability can be imbedded into existing CCTV networks, making its operation impossible to detect. Furthermore, it is passive in its operation. It requires no participation or consent from its targets – it is 'non-intrusive, contact-free process' (Woodward et al. 2003: 7). Its application is flexible. It can as easily be used by a supermarket to monitor potential shoplifters (as was proposed and later abandoned, by the Borders bookstore), by casinos to track potential fraudsters, by law enforcement to monitor spectators at a Super Bowl match (as was done in Tampa, Florida), or used for identifying 'terrorists' at airports (as is currently in operation at various US airports). However, most important of all is the obscurity of its operation.

Most of the software algorithms at the heart of facial recognition systems (and other information technology products) are propriety software objects. Thus, it is very difficult to get access to them for inspection and scrutiny. More specifically, however, even if you can go through the code line by line, it is impossible to inspect that code *in operation*, as it becomes implemented through multiple layers of translation for its execution. At the most basic level we have electric currents flowing through silicon chips, at the highest level we have programme instructions, yet it is almost impossible to trace the connection between these *as it is being executed*. Thus, it is virtually impossible to know if the code you inspected is the code being executed, when executed. In short, *software algorithms are operationally obscure*.

It is our argument that the opaque and 'silent' nature of digital technology makes it particularly difficult for society to scrutinise it. Furthermore, this inability to scrutinise creates unprecedented opportunities for this silent and 'invisible' micro-politics to become pervasive (Graham and Wood 2003). Thus, a profound sort of micro-politics can emerge as these opaque (closed) algorithms become enclosed in the social-technical infrastructure of everyday life. We tend to have extensive community consultation and impact studies when we build a new motorway. However, we tend not to do this when we install CCTV in public places or when we install facial recognition systems in public spaces such as airports, shopping malls, etc. To put is simply: most informed people tend to understand the cost (economic, personal, social, environmental) of more transparent technologies such as a motorway, or a motorcar, or maybe even cloning. However, we would argue that they do not often understand the 'cost' of the more opaque information technologies that increasingly pervade our everyday life. We will aim to disclose this

Table 1. Opaque versus transparent technology

Opaque technology is:	*Transparent* technology is:
Embedded/hidden	On the 'surface'/conspicuous
Passive operation (limited user involvement, often automatic)	Active operation (fair user involvement, often manual)
Application flexibility (open ended)	Application stability (firm)
Obscure in its operation/outcome	Transparent in its operation/outcome
Mobile (*soft*-ware)	Located (*hard*-ware)

in the case of facial recognition systems below. Before we do this we want to give an account of what we mean by this 'disclosure' of disclosive ethics.

Disclosive ethics as the 'other' side of politics

Ethics is always and already the 'other' side of politics (Critchley 1999). When we use the term 'politics' (with a small 'p') – as indicated above – we refer to the *actual operation of power* in serving or enclosing particular interests, and not others. For politics to function as politics it seeks closure – one could say 'enrolment' in the actor network theory language. Decisions (and technologies) need to be *made* and programmes (and technologies) need to be *implemented*. Without closure politics cannot be effective as a programme of action and change. Obviously, if the interests of the many are included – in the *en*closure as it were – then we might say that it is a 'good' politics (such as democracy). If the interests of only a few are included we might say it is a 'bad' politics (such as totalitarianism). Nevertheless, all political events of enclosing are violent as they always include and exclude as their condition of operation.

It is the excluded – the *other* on the 'outside' as it were – that is the concern of ethics. Thus, every political action has, always and immediately, tied to its very operation an ethical question or concern – it is the other side of politics. When making this claim it is clear that for us ethics (with a small 'e') is not ethical theory or moral reasoning about how we ought live (Caputo 1993). It is rather the question of the actual operation of closure in which the interests of some become excluded as an implicit part of the material operation of power – in plans, programmes, technologies and the like. More particularly, we are concerned with the way in which the interest of some become excluded through the operation of closure as an implicit and essential part of the *design* of information technology and its *operation* in social-technical networks.

As those concerned with ethics, we can see the operation of this 'closure' or 'enclosure' in many related ways. We can see it operating as already 'closed' from the start – where the voices (or interests) of some are shut out from the design process and use context from the start. We can also see it as an ongoing operation of 'closing' – where the possibility for suggesting or requesting alternatives are progressively excluded. We can also see it as an ongoing operation of 'enclosing' – where the design decisions become progressively 'black-boxed' so as to be inaccessible for further scrutiny. And finally, we can see it as 'enclosed' in as much as the artefacts become subsumed into larger socio-technical networks from which it becomes difficult to 'unentangle' or scrutinise. Fundamental to all these senses of closure is "the event of closure [as] a delimitation which shows the double appartenance of an inside and an outside..." (Critchley 1999: 63).

We need to acknowledge that politics – or the operation of closure – is fundamental to the ongoing production of social order. Decisions have to be made, technologies have to be designed and implemented, as part of the ongoing ordering of society. Agendas cannot be kept open forever, designs cannot be discussed and considered indefinitely. Thus, we are not suggesting an end to politics as the operation of closure. Closure is a pragmatic condition for life. Equally, we are not arguing that the question of ethics can, and ought to be, 'divorced' from politics. Ethics cannot escape politics. The concern of ethics is always and already also a political concern. To choose, propose or argue for certain values – such as justice, autonomy, democracy and privacy as suggested by Brey (2000) – is already a political act of closure. We may all agree with these values as they might seem to serve our interests, or not. Nevertheless, one could argue that they are very anthropocentric and potentially excludes the claims of many others – animals, nature, the environment, things, etc.

If ethics cannot escape politics then it is equally true that politics cannot escape ethics. This is our starting point – a powerful one in our view. The design or use of information technology is not morally wrong as such. The moral wrongdoing is rather the nondisclosure of the closure or the operation of politics as if ethics does not matter – whether it is intended or not. We know that power is most effective when it hides itself (Foucault 1975). Thus, power has a very good reason to seek and maintain nondisclosure. Disclosive ethics takes as its moral imperative the disclosure of this nondisclosure – the presumption that politics can operate without regard to ethics – as well as the *dis*closure of all attempts at closing or enclosing that are implicitly part of the design and use of information technology in the pursuit of social order.

Obviously at a curtain level design is rather a pragmatic question. However, it is our contention that many seemingly pragmatic or technical decisions may have very important and profound consequences for those excluded – as we will show below. This is the important task of disclosive ethics. Not merely to look at this or that artefact but to trace all the moral implications (of closure) from what seems to be simple pragmatic or technical decisions – at the level of code, algorithms, and the like – through to social practices, and ultimately, to the production of

particular social orders, rather than others. For disclosive ethics it is the way in which these seemingly pragmatic attempts at closing and enclosing connect together to deliver particular social orders that excludes some and not others – irrespective of whether this was intended by the designers, or not. Indeed it will be our argument that in the design of complex socio-technical networks these exclusionary or enclosing possibilities often do not surface as a consideration when making this or that particular design decision. Furthermore, these exclusionary possibilities often emerge as a systemic effect or outcome with no particular 'author' in charge of the script as such. Indeed this is what we intend to show in the disclosing of facial recognition systems below. In summary, disclosive ethics operates with two principles:

(a) To disclose the nondisclosure of politics by claiming a place for ethics as being always and immediately present in every actual operation of power
(b) To trace and disclose the intentional or unintentional *en*closure of values and interests from every minute technical detail through to social practices and complex social-technical networks.

We will now turn our attention to a disclosive analysis of facial recognition systems in order to disclose its politics and the way in which these may emerge in the social practices of securing identity.

Disclosing facial recognition systems

Getting a digital face: the facial recognition system

Figure 1 below depicts the typical way that a facial recognition system (FRS) system can be made operational.

The first step is the capturing of a face image. This would normally be done using a still or video camera. As such it can be incorporated into existing 'passive' CCTV systems. However, locating a face image in the field of vision is not a trivial matter at all. The effectiveness of the whole system is dependent on the quality of the captured face image. The face image is passed to the recognition software for recognition (identification or verification). This would normally involve a number of steps such as normalising the face image and then creating a 'template' of 'print' to be compared to those in the database. If there is a 'match' then an alarm would solicit an operator's attention to verify the match and initiate the appropriate action. The match can either be a true match which would lead to investigative action or it might be a 'false positive' which means the recognition algorithm made a mistake and the alarm would be cancelled. Each element of the system can be located at different locations within a network, making it easy for a single operator to respond to a variety of systems.

For our analysis we want to concentrate on steps two and three of the system. We want to scrutinise the FR algorithms, the image database (also called the gallery) and the operators. At each of these points important decisions are made which may have important ethical and political implications.

Facial recognition algorithms and reduction

Research in software algorithms for facial recognition has been ongoing for the last 30 years or so (Gross et al. 2001a). However, advances in information technology and statistical methods have given impetus to this development with seemingly excellent recognition results and low error rates – at least in ideal laboratory conditions. It is possible to identify

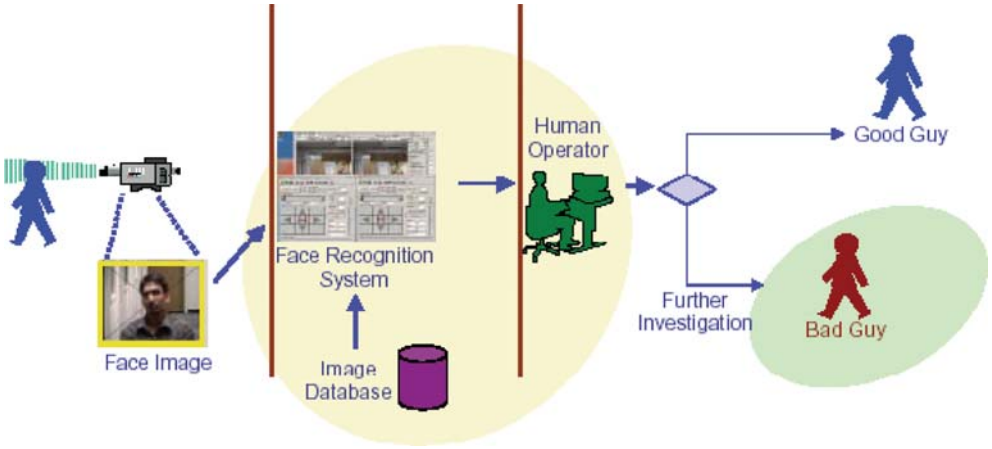

Figure 1. Overview of FRS (Source: Face Recognition, Vendor Test 2002).

two main categories of algorithms according to Gross et al. (2001a):

The image template algorithms. These algorithms use a template-based method to calculate the correlation between a face and one or more standard templates to estimate the face identity. These standard templates tend to capture the global features of a gallery of face images. Thus, the individual face identity is the difference between (or deviation from) the general or 'standard' face. This is an intuitive approach since we, as humans tend to look for distinctive features (or differences from the general) when we identify individuals. Some of the methods used are: Support Vector Machines (SVM), Principal Component Analysis (PCA), Neural Networks, Kernel Methods, etc. The most commercially known template based algorithm is the MIT Bayesian Eigenface technique, which has been developed with the PCA method. During various tests conducted in 1996, its performance was consistently near the top compared to other available at the time.

The geometry feature-based algorithms. These methods capture the local facial features and their geometric relationships. They often locate anchor points at key facial features (eyes, nose, mouth, etc.), connect these points to form a net and then measure the distances and angles of the net to create a unique face 'print'. The most often cited of these is the technique known as Local Feature Analysis (LFA), which is used in the Identix (formerly known as Visionics) face recognition system called FaceIt. The LFA method, in contrast to the PCA technique, is less sensitive to variations in lighting, skin tone, eye glasses, facial expression, hair style, and individual's pose up to 35 degrees.

The commonality in both of these groups of techniques is the issue of *reduction*. In order to be efficient in processing and storage the actual face image gets reduced to a numerical representation (as small as 84 bytes or 84 individual characters in the case of FaceIt). With this reduction certain information is disregarded (as incidental or irrelevant) at the expense of others. It is here that we need to focus our analysis. What is the consequences of the process of reduction. It would be best to understand this through some detailed study of the logic and operation of these algorithms in diverse settings with diverse databases. This has not yet being done (not even in the Facial Recognition Vendor Tests of 2002 (FRVT 2002) which has been the most comprehensive thus far). Nevertheless, with our limited knowledge we can make some logical conclusions and then see how these may play out in the FRVT 2002 evaluations. How will the reduction, effect the performance of these algorithms?

- *Template based algorithms*. In these algorithms certain biases become built into the standard template. It obviously depends on the gallery used to create the standard template as well as the range of potential variations within a population. For example, because minorities tend to deviate the most from the standard template they might become easier to recognise.
- *Feature based algorithms*. These algorithms do not have an initial bias. However, because of the reduction the 'face prints' generated are in close proximity to each other. Thus, as the gallery database increases more and more face prints are generated in ever diminishing proximity, thereby making the discrimination required for the recognition task more difficult. Therefore, the operation of the system deteriorates rapidly as the database increases (this is also true for template based algorithms). It also makes the system dependent on good quality face images. The implication of this is that the system will operate at its best with a small database and good quality face capture, such as an operator assisted face capture (reintroducing the operator bias). In addition to this, it will tend to be better at identifying those that are more distinctive, or less similar, to those already in the database (such as minorities).

Thus, in both cases we would expect some form of bias to emerge as a result of the reduction. Is this conclusion borne out by the performance of these algorithms in the FRVT? Let us now consider the results of these evaluations.

The evaluations: reduction, operation and error

The most significant evaluation of FRSs happened with the Facial Recognition Vendor Tests of 2002 (Phillips et al. 2003). These test were independent tests sponsored by a host of organizations such as Defense Advanced Research Projects Agency (DARPA), the Department of State and the Federal Bureau of Investigation. This evaluation followed in the footsteps of the earlier FRVT of 2000 and the FERET evaluations of 1994, 95 and 96. In the FRVT 2002 ten FRS vendors participated in the evaluations. The FRVT of 2002 were more significant than any of the previous evaluations because of:

- The use of a large database (37,437 individuals)
- The use of a medium size database of outdoor and video images
- Some attention given to demographics

The large database (referred to as the HCInt data set) is a subset of a much larger database which was

provided by the Visa Services Directorate, Bureau of Consular Affairs of the U.S. Department of State. The HCInt data set consisted of 121,589 images of 37,437 individuals with at least three images of each person. All individuals were from the Mexican non-immigrant visa archive. The images were typical visa application type photographs with a universally uniform background, all gathered in a relatively consistent manner.

The medium size database consisted of a number outdoor and video images from various sources. Figure 2 below gives an indication of the images in the database. The top row contains images taken indoors and the bottom contains outdoor images taken on the same day. Notice the quality of the outdoor images. The face is consistently located in the frame and similar in orientation to the indoor images.

For the identification task an image of an unknown person is provided to a system (assumed to be in the database). The system then compares the unknown image (called the probe image) to the database of known people. The results of this comparison are then presented by the system, to an operator, in a ranked listing of the top *n* 'candidates' (referred to as the 'rank', typically anywhere from 1 to 50). If the correct image is somewhere in the top *n*, then the system is considered to have performed the identification task correctly. Figure 3 below indicates the performance at rank 1, 10 and 50 for the three top performers in the evaluation.

Figure 2. Indoor and outdoor images from the medium data base. (from FRVT2002 report, p. 16).

With the very good images from the large database (37,437 images) the identification performance of the best system at rank one is 73% at a false accept rate of 1%. There is a tradeoff between the recognition rates and the level of 'false accepts' (incorrect identification) one is prepared to accept, the false accept rate. If you are prepared to accept a higher false accept rate then the recognition performance can go up. However, this will give you more cases of false identification to deal with. This rate is normally a threshold parameter that can be set by the operators of the system.

What are the factors that can detract from this 'ideal' performance? There might be many. The FRVT 2002 considered three of the most important ones:

- Indoor versus outdoor images
- The time delay between the database image and the probe image
- The size of the database

The identification performance drops dramatically when outdoor images are used – in spite of the fact that they can be judge as relatively good – as indicated above. One would not expect a typical video camera to get this quality of image all the time. For the best systems the recognition rate for faces captured *outdoors* (i.e. less than ideal circumstances) was only 50% at a false accept rate of 1%. Thus, as the report concluded: "face recognition from outdoor imagery remains a research challenge area." The main reason for this problem is that the algorithm cannot distinguish between the change in tone, at the

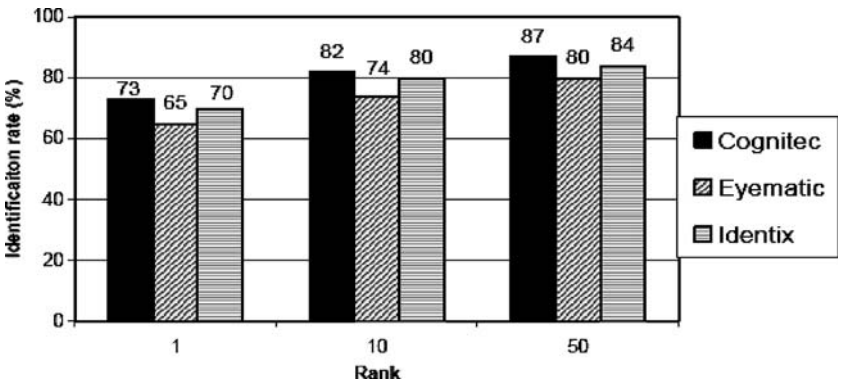

Figure 3. Performance at rank 1, 10 and 50 for the three top performers in the evaluation (from FRVT 2002, Overview and Summary, p. 9).

pixel level, caused by a relatively dark shadow, versus such a change caused by a facial feature. As such it starts to code shadows as facial features. The impact of this on the identification may be severe if it happens to be in certain key areas of the face.

As one would expect, the identification performance also decreases as time laps increases between the acquisition of the database image and the newly captured probe image presented to a system. FRVT 2002 found that for the top systems, performance degraded at approximately 5% points per year. It is not unusual of the security establishment to have a relatively old photograph of a suspect. Thus, a two year old photograph will take 10% off the identification performance. A study by the US National Institute of Standards and Technology found that two sets of mugshots taken 18 months apart produced a recognition rate of only 57% (Brooks 2002). Gross et al. (2001: 17) found an even more dramatic deterioration. In their evaluation, the performance dropped by 20% in recognition rate for images just two weeks apart. Obviously these evaluations are not directly comparable. Nevertheless, there is a clear indication that there may be a significant deterioration when there is a time gap between the database image and the probe image.

What about the size of the database? For the best system, "the top-rank identification rate was 85% on a database of 800 people, 83% on a database of 1,600, and 73% on a database of 37,437. For every doubling of database size, performance decreases by two to three overall percentage points" (Phillips et al. 2003: 21). What would this mean for extremely large databases? For example, the UK fingerprint database consists of approximately 5.5 million records. If one had a similar size 'mugshot' database how will the algorithms perform in identifying a probe image in that database? If one takes the decrease to be 2.5% for every doubling of the database, and use 73% at 37,437 as the baseline, then one would expect the identification performance to be approximately 55% in ideal conditions and as low as 32% in less than ideal conditions.

To conclude this discussion we can imagine a very plausible scenario where we have a large database, less than ideal image due to factors such as variable illumination, outdoor conditions, poor camera angle, etc. and the probe image is relatively old, a year or two. Under these conditions the probability to be recognized is very low, unless one sets the false accept rate to a much higher level, which means that there is a risk that a high number of individual may be subjected to scrutiny for the sake of a few potential identifications. What will be the implications of this for practice? We will take up this point again below. Obviously, we do not know how these factors would act together and they are not necessarily cumulative. Nevertheless, it seems reasonable to believe that there will be some interaction that would lead to some cumulative affect.

Such a conclusion can make sense of the Tampa Police Department case reported by ACLU (Stanley and Steinhardt 2002) as well as the Palm Beach International Airport also reported by the ACLU. In the Tampa case the system was abandoned because of all the false positive alarms it generated. As far as it could be ascertained it did not make one single positive identification. In the Palm Beach Airport case the system achieved a mere 47% correct identifications of a group of 15 volunteers using a database of 250 images (Brooks 2002). In Newham, UK, the police admitted that the FaceIt system had, in its two years of operation, not made a single positive identification, in spite of working with a small database. One could argue that there might not have been the potential for a match to be made as none of the individual in the database actually appeared in the street. Nevertheless, the system could not identify a *Guardian* journalist, placed in the database, that intentionally presented himself in the two zones covered by the system (Meek 2002). These cases indicate the complexity of real world scenarios. We now want to move to the focal concern of this paper namely the question of biases in the algorithms themselves.

Reduction and biased code

The most surprising outcome – for those involved – of the FRVT 2002 is the realization that the algorithms displayed particular identification biases. First, recognition rates for males were higher than females. For the top systems, identification rates for males were 6–9% points higher than that of females. For the best system, identification performance on males was 78% and for females was 79%. Second, recognition rates for older people were higher than younger people. For 18–22 year olds the average identification rate for the top systems was 62%, and for 38–42 year olds was 74%. For every 10 years increase in age, on average performance increases approximately 5% through to age 63. Unfortunately, they could not check race as the large data set consisted of mostly Mexican non-immigrant visa applicants. However, research by Givens et al. (2003), using PCA algorithms, has confirmed the biases in the FRVT 2002 (except for the gender bias) and also found a significant race bias. This was confirmed using balanced databases and controlling for other factors. They concluded that: "Asians are easier [to

recognize] than whites, African-Americans are easier than whites, other race members are easier than whites, old people are easier than young people, other skin people are easier to recognize than clear skin people...'' (p. 8). Their results are indicated in Figure 4 below.

These results were also found in another context by Furl, Phillips and O'Toole (2002) in their study of recognition performance by 13 different algorithms. One can legitimately ask whether these differences, probably in the order of 5–10%, really makes a difference? Are they not rather trivial? We would argue that taken by themselves they may seem rather trivial. However, as we argued earlier on, it is when these trivial differences become incorporated into a network of practices that they may become extremely important. This is what we now want to explore: the politics of the digital face as it becomes imbedded in practices.

Closure and the ethics of the digital face

FRSs: efficient, effective and neutral

Many security analysts see FRSs as the ideal biometric to deal with the new emerging security environment (post 11 September). They claim that it is efficient (FaceIt only requires a single 733 Mhz Pentium PC to run) and effective, often quoting close to 80% recognition rates from the FRVT 2002 evaluation while leaving out of the discussion issues of the quality of the images used in the FRVT, the size of the database, the elapsed time between database image and probe image, etc. But most of all they claim that these systems ''performs equally well on all races and both genders. Does not matter if population is homogeneous or heterogeneous in facial appearance'' (Faceit technical specification[1]). This claim is not only made by the suppliers of FRSs such as Identix and Imagis Technologies. It is also echoed in various security forums: ''Face recognition is completely oblivious to differences in appearance as a result of race or gender differences and is a highly robust Biometrics''[2] Even the critical scholar Gary Marx (1995: 238) argued that algorithmic surveillance provides the possibility of eliminating discrimination. The question is not whether these claims are correct or not. One could argue that in a certain sense they are correct. The significance of these claims is the way they *frame* the technology. It presents the technology itself as neutral and unproblematic. More than this it presents the technology as a solution to the problem of terrorism. Atick of Identix claimed, in the wake of the 9/11 attacks, that with FaceIt the US has the ''ability to turn all of these cameras around the country into a national shield'' (O'Harrow 2001). He might argue that in the face of terrorism 'minor' injustices (biases in the algorithms) and loss of privacy is a small price to pay for security. This may be so, although we would disagree.

Nevertheless, our main concern is that these arguments present the technical artefacts in isolation with disregard to the socio-technical networks within which they will become enclosed. As argued above, it is not just the micro-politics of the artefact that is the issue. It is how these become multiplied and magnified as they become tied to other social practices that is of significance. We need to disclose the 'network effects', as it were, of the micro-politics of artefacts. This is especially so for opaque digital technology. There is every reason to believe that the silent and non-invasiveness of FRSs make it highly desirable as a biometric for digital surveillance. It is therefore important that this technology becomes disclosed for its potential politics in the socio-technical network of digital surveillance. Thus, not just as isolated software objects as was done in the FRVTs but in its multiplicity of implementations and practices. We would claim it is here where the seemingly trivial exclusions may become very important as they become incorporated into actual practices.

FRSs and the production of suspects

There is an urgent need for an in-depth study of FRSs in practice (as has been done with CCTV by Norris and Armstrong (1999) and others). However, since we currently only have a limited number of systems in operation and due to the sensitivity of these implementations it is unlikely that we would be able to do so in the near future. Thus, in the face of this limitation, we propose to outline what we consider to be a highly probable scenario of how these digital closures may become incorporated into other practices that would render these seemingly trivial biases significant.

Based on the FRVT of 2002 we know that, although FRSs have the capability to achieve a 70–85% accuracy rate, this is only in ideal circumstances. The system's performance degrades significantly in an uncontrolled 'face-in-the-crowd' environment, with a large database, and where there is an elapsed time between the database image and the probe image. This would seem to us to be a usual rather than an unusual situation. What will happen if the system's performance degrades under these rather usual conditions?

[1] http://www.identix.com/newsroom/news_biometrics_face_acc.html

[2] http://www.ats-computers.com/biometrics/face.html http://www.biocom.tv/BIOMETRICS_types.htm

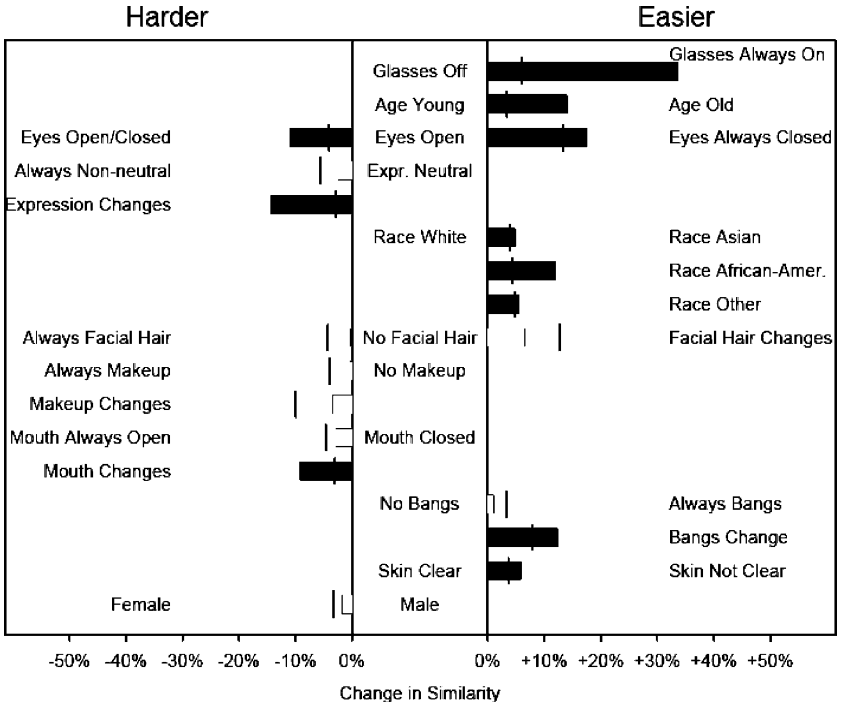

Figure 4. From Givens et al. (2003) indicating which factor make it harder or easier to correctly identify a probe image presented to a system.

We would propose that two possibilities are most likely. First, it is possible that the operators will become so used to false positives that they will start to treat all alarms as false positives thereby rendering the system useless. Alternatively, they may deal with it by increasing the identification threshold (requesting the system to reduce the number of false positives). This will obviously also increase the false negatives, thereby raising all sorts of questions about the value of the system into question. However, more important to us, with an increased threshold small differences in identifiability (the biases outlined above) will mean that those that are easier to identify by the algorithms (African-Americans, Asians, dark skinned persons and older people) will have a greater probability of being scrutinised. If the alarm is an actual positive recognition then one could argue that nothing is lost. However, it also means that these groups would be subjected to a higher probability of scrutiny as false positives, i.e. mistaken identity. Moreover, we would propose that this scrutiny will be more intense as it would be based on the assumption that the system is working at a higher level and therefore would be more accurate. In such a case existing biases, against the usual suspects (such as minorities), will tend to come into play (Norris and Armstrong 1999). The operators may even override their own judgements as they may think that the system under such high conditions of operation must 'see something' that they do not. This is highly likely as humans are not generally very good at facial recognition in pressurised situations as was indicated in a study by Kemp et al. (1997). Thus, under these conditions the bias group (African-Americans, Asians, dark skinned persons and older people) may be subjected to disproportionate scrutiny, thereby creating a new type of 'digital divide' (Jupp in Graham and Wood, 2003: 234).

How likely is this scenario? We believe it to be more likely than we presume. We have only the following anecdotal evidence reported in the *Discover Magazine* of an installation at the Fresno Yosemite International Airport to suggest:

> "[The system] generates about one false positive for every 750 passengers scanned, says Pelco vice president Ron Cadle. Shortly after the system was installed, a man *who looked as if he might be from the Middle East* set the system off. "The gentleman was detained by the FBI, and he ended up spending the night," says Cadle. "We put him up in a hotel, and he caught his flight the next day." (Garpinkle 2002, p. 19 – *emphasis added*)

To produce only one false positive per 700 passengers the system had to operate with a very restricted false positive rate, thereby suggesting that an alarm must 'mean something'. Notice that one of the false positives was a man supposedly from 'Middle Eastern'

origin. The individual was detained and questioned by the FBI because he "looked as if he might be from the Middle East" in spite of the fact that he was obviously a false positive. There could be many explanations for this action. Nevertheless, it is likely that they may have decided to detain him 'just in case' the system saw something they did not see. This is likely in a situation where a human operator must make a decision. We know from research that humans find it very difficult to identify individual from other ethnic groups (Kemp et al. 1997), exactly the group that we would expect to emerge as likely false positives. In these moments of uncertainty, the FRSs may be taken as more authoritative than the humans involved. This case clearly demonstrates the scenario we outline above. Our disclosive analysis has demonstrated that seemingly trivial differences in recognition rates, within the algorithm, can indeed have important political (ethical) implications for some when it becomes incorporated into a whole set of socio-technical surveillance practices.

One might imagine that in an environment where there is an acute sense of vulnerability it would not be unreasonable to store these false positives in a database 'just in case' (Lyon 2001, 2002). These false positive may then become targets for further scrutiny. Why? Just because they have features that make them more distinctive. We are not saying that this will happen. We are merely trying to indicate how seemingly trivial 'technical issues' can add up to strong political ideologies at the expense of some for the sake of others. This is the issue of the politics – and ethics – of FRSs. This is particularly dangerous politics in the case of silent and opaque technologies such as FRSs. Obviously more in-depth study of actual installations are required.

There is no doubt in our minds that facial biometric is a very important part of the future security infrastructure. Kopel and Krause (2003) reports that: "As of June 2001 the Departments of Justice and Defence had given about $21.3 million and $24.7 million, respectively, to the research and development of FRSs." Its efficiency, ease of implementation and invisible nature make it and ideal biometric. We believe, we have demonstrated that there are many aspects of this opaque technology that still needs to be disclosed (see Agre 2003 for more indications of what might be disclosed).

Nevertheless, this disclosive analysis of facial recognition systems is not complete. We have not looked at those that have been excluded from the start. For example, the fact that most of the research in FRSs are sponsored by US government agencies, who has been excluded through this mechanism? We have not considered the way in which equally valid other alternatives have become progressively excluded. What other ways of securing is possible? More importantly, we have not disclosed ourselves as those doing the disclosing. How does this analysis itself enclose? Indeed, disclosive ethics is an infinite task. We believe it is worth doing even if there is no clear guidelines and no clear end. This is in our view not a flaw but rather its strength. It expects every closure to be disclosed irrespective of where it emanates from – that is why it is disclosive.

References

M. Akrich. The De-scription of Technical Objects. In W.E. Bijker and J. Law, editors, *Shaping Technology/Building Society*, pp. 205–224. MIT Press, Cambridege, 1992.

P.E. Agre. Your Face is Not a Bar Code: Arguments Against Face Recognition in Public Places, [Online], 2003. Available: http://dlis.gseis.ucla.edu/pagre, [2003, May 25].

P. Agre and C. Mailloux. Social Choice about Privacy: Intelligent Vehicle-Highway Systems in the United States. In B. Friedman, editor, *Human Values and Design of Computer Technology*, Cambridge University Press, Cambridge, 1997.

Brooks, Michael. "Face-off", New Scientist, Vol. 175, Issue 2399, 9/7/2002.

Philip. Brey. Disclosive Computer Ethics. *Computers and Society*, 30(4): 10–16, 2000.

Michel. Callon. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. In John. Law, editor, *Power, Action and Belief*, pp. 196–233. Routledge & Kegan Paul, London, 1986.

John D. Caputo, *Against Ethics*. Indiana University Press, Indianapolis, 1993.

Simon. Critchley, *The Ethics of Deconstruction: Derrida and Levinas*. 2nd edn. Edinburgh University Press, Edinburgh, 1999.

Andrew. Feenberg, *Questioning Technology*. London and New York, Routledge, 1999.

N. Furl, J.P. Phillips and A.J. O'Toole. Face Recognition Algorithms and the Other-race Effect: Computational Mechanisms for a Developmental Contact Hypothesis. *Cognitive Science*, 26: 797–815, 2002.

Facial Recognition Vendor Test 2002 (FRTV2002), [Online], Available: http://www.frvt.org/FRVT2000/default.htm, [2003, Aug.1].

M. Foucault, *Discipline and Punish, The Birth of the Prison*. Penguin Books Ltd., London, UK, 1975.

S. Garpinkle. Don't Count on Face-recognition Technology to Catch Terrorists. *Discover*, 23(9): 17–20, 2002.

G. Givens, J.R. Beveridge, B.A. Draper, and D. Bolme. A statistical Assessment of Subject Factors in the PCA Recognition of Human Faces, [Online], 2003. Available: http://www.cs.colostate.edu/evalfacerec/papers/csusacv03.pdf,[2003, July.10].

S. Graham and D. Wood. Digitizing Surveillance: Categorization, Space and Inequality. *Critical Social Policy*, 20(2): 227–248, 2003.

R. Gross, J. Shi, and J.F. Cohn. Quo vadis Face Recognition?, [Online], 2001a. Available: http://dagwood.vsam.ri.cmu.edu/ralph/Publications/QuoVadisFR.pdf, [2003, July 10].

L.D. Introna. Oppression, Resistance and Information Technology: Some Thoughts on Design and Values, Design for Values: Ethical, Social and Political Dimensions of Information Technology workshop sponsored by the NSF DIMACS held at Princeton University, USA, February 27-March 1, 1998.

L.D. Introna and H. Nissenbaum. The Internet as a Democratic Medium: Why the Politics of Search Engines Matters. *Information Society*, 16(3): 169–185, 2000.

F. Kafka, *The trial*. Penguin Books Ltd., London, England, 1925.

R. Kemp, N. Towell and G. Pike. When Seeing Should not be Believing: Photographs, Credit Cards and Fraud. *Applied Cognitive Psychology*, 11: 211–222, 1997.

D. Kopel and M. Krause. Face the FactsFacial recognition technology's troubled past – and troubling future. [Online] 2003. Available: http://www.reason.com/0210/fe.dk.face.shtml, [2003, Nov. 18].

B. Latour. Technology is Society Made Durable. In J. Law, editor, *A Sociology of Monsters: Essays on Power, Technology and Domination*, pp. 103–131. Routledge, London, 1991.

B. Latour. Where are the Missing Masses? The Sociology of a Few Mundane Artefacts. In W.L.J. Bijker, editors, *Shaping Technology/Building Society*, pp. 225–258. MIT Press, London, 1992.

John. Law, *The Sociology of Monsters: Essays on Power, Technology and Domination*. Routledge, London, 1991.

D. Lyon, *Surveillance Society, Monitoring Everyday Life*. Open University Press, Philadelphia, USA, 2001.

D. Lyon. Surveillance After September 11, 2001, [Online], 2002. Available: http://www.fine.lett.hiroshima-u.ac.jp/lyon/lyon2.html, [2003, July.10].

G.T. Marx. The Engineering of Social Control: The search for the silver Bullet. In J. Hagen and R. Peterson, editors, *Grime and Inequality*, pp. 225–46. Stanford University Press, Stanford, CA, 1995.

J. Meek. Robo cop: Some of Britain's 2.5 million CCTV cameras are being hooked up to a facial recognition system designed to identify known criminals. But does it work, Guardian, June 13, 2002.

C. Norris and G. Armstrong, *The Maximum Surveillance Society, The Rise of CCTV*. Berg, New York, USA, 1999.

R. O'Harrow Jr. Facial Recognition System Considered For US Airports, *Washington Post*, Monday, September 24, 2001, Page A14.

P. Phillips, P. Grother, R. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone. Face Recognition Vendor Test 2002: Overview and Summary, [Online], 2003. Available: http://www.biometricsinstitute.org/bi/ FaceRecognitionVendorTest2002.pdf, [2003, May 30].

J. Stanley and B. Steinhardt. Drawing a Blank: The Failure of Facial Recognition Technology in Tampa, Florida, ACLU Special Report, 2002.

Landon. Winner. Do Artefacts Have Politics. *Daedalus*, 109: 121–136, 1980.

J. Woodward, C. Horn, J. Gatune and A. Thomas. Biometrics: A Look at Facal Recognition, Documented Briefing prepared for the Virginia State Crime Commission, 2003 (available at http://www.rand.org).