SUPERIDENTITY: An Exploration of Identity across online and offline contexts

Sarah Stevenage
Sue Black
Sadie Creese
Richard Guest
Bill Pike
Steve Saxby
Danaë Stanton Fraser
Monica Whitty
Chris Bevan
Lia Emanuel
Rachel Fletcher
Hongmei He
Duncan Hodges
Sue Jamison-Powell
Oriana Love
Greg Neil
Jean Scholtz

Sept 2012

Table of Contents

1. The SuperIdentity Team

The Super-Identity project is an ambitious proposal covering a range of disciplines. This annual report outlines the knowledge and experience of the Investigators, and the progress made within Year 1 of the Project.

Anatomical and Behavioural Indicators of Identity: Offline World



Expertise is provided by Professor Sue Black (Dundee), Dr Richard Guest (Kent) and Dr Sarah Stevenage (Southampton). Together, they bring considerable experience in anatomical and biometric measures of identity in the real world environment. Professor Black is the most experienced forensic anthropologist in the UK advising on issues of identification both at home and overseas. Dr Guest brings expertise in the field of automated biometric systems most notably in the areas of handwriting and dynamic signature verification, biometric image analysis, classification architectures and system interaction. Finally Dr Stevenage brings a cognitive psychology perspective on the human capacity to identify individuals from a range of static and dynamic cues in the real world including the face, voice, and gait. Together, these Investigators hold grants totalling nearly £5million from EPSRC, EU and other national and international funding bodies including government and industry. In addition, the Investigators provide representation to policy makers at the highest level including UK Government, Interpol, and International Standards (BSI and ISO).

Novel Behavioural Indicators of identity: Cyber World


Expertise is provided by Professor Monica Whitty (Leicester) and Professor Danaë Stanton Fraser (Bath). Professor Whitty's main area of expertise is cyberpsychology, with a focus on the capacity to self-present either truthfully or untruthfully through cyber behaviour. Recent work explores online relationships, internet infidelity, representation of self online, use of the internet by married couples, cyberstalking, Internet surveillance, deception across different mediums, engaging in symbolic taboo activities in video games, and online scams. She has been the PI on several grants notably on online surveillance and privacy; and deception across different modes of communication. Currently she is the PI on an ESRC funded project on the online romance scam. Professor Stanton Fraser's area of expertise is human-computer interaction, with a focus on exploration of adults and young people's interactions with technology. She has been funded by numerous research council, business/industry and charity awards. She was CI on the

EPSRC ‘Cityware’ project exploring trust relationships in the design of mobile and pervasive applications; and PI on the DTI/EPSRC ‘Participate’ project exploring pervasive computing for mass participation in environmental monitoring.

Digital Security, Modelling and Data Visualisation


Expertise is provided by Professor Sadie Creese (Oxford) and Dr Bill Pike (PNNL – US). Professor Creese is Professor of CyberSecurity at Oxford University, and is based in the Department of Computer Science. She is recipient of an IBM Faculty Award (2009) and is a member of various advisory groups with concerns spanning ‘Global Uncertainties’, the International Systems Security Association UK, and Cloud Security. She is PI on 3 collaborative projects funded by EPSRC. Dr Pike is a Senior Research Scientist in visual analytics, and research coordinator for the National Visualization and Analytics Center at PNNL. In conjunction with both government and industrial partners, he leads work on behavioural modelling of actors on a computer network for anomaly detection, the creation of temporal visualization techniques for pattern discovery in communications activity, interactive decision support capabilities for emergency management, and online visualization tools for the personalized display of social network data. He served as Chair of the 2010 and 2011 IEEE Conferences on Visual Analytics Science and Technology.

Legal and Ethical Representation

Expertise is provided by Professor Steve Saxby (Southampton). Professor Saxby is Director of the Institute for Law and the Web and is Professor of IT Law and Public Policy. He is co-founder of the International Association of IT Lawyers and the LSPI conference. He formerly served on the Legal Advisory Board of the European Commission, and the Intellectual Property Committee of the British Computer Society. He has been a Consultant to JISC; Ordnance Survey; Netherlands Council for Geographic Information; Countryside Agency, and Southampton City Council. Notable recent activities include the 2010 'WeGov' project (Where e-Government meets the e-Society) and legal consultation to the GeoData Insititute in their audit of data policy for the Crown Estate Office.

Contact Us:

By Mail:

SuperIdentity Principal Investigator:

Dr Sarah Stevenage

Psychology

University of Southampton

Highfield, Southampton,

Hampshire

SO17 1BJ

By Telephone:

SuperIdentity Administrator: Mrs Barbara Seiter

Tel: 02380 595578

By Email:

Superidentity@soton.ac.uk

Our Website:

www.superidentity.org

www.soton.ac.uk/superidentity

2. Executive Summary

During Year 1 of the SuperIdentity project, we have addressed questions regarding the way in which individuals represent themselves both offline and online. Our findings suggest that context is paramount. A simple distinction between offline and online identity is not as important as one may imagine, either in terms of self-presentation or in terms of deception. Instead, the intention that people have in each space is of value as this shapes the aspects of themselves that they show, and that they leak.

Across a series of biometrics, assessment of recognition shows the importance of the source of the recognition (expert versus novice?), the conditions of recognition (optimal versus poor), and the measure being recognised (stable versus changeable). Both human recognition and automated recognition will suffer to differing degrees from these limitations. However, application of this framework allows us to determine the level of confidence that we might have in an identity judgement from each biometric.

Across a series of cybermetrics, a similar picture emerges. Here, however, our science is younger, and work is still required to establish the reliability of an identification from a given cybermetric measure. We have focussed in particular on those cybermetrics which allow identity to be matched from online to offline contexts, and these combinations are articulated in our SuperIdentity model.

The SuperIdentity model itself operationalizes a series of theoretically driven, and empirically demonstrated, linkages between metrics in the real and the cyber world. Together, these reveal identity. Each link carries a level of confidence. In combination, an identification can then be judged to be more or less robust. Through interviews with identified users from law enforcement, intelligence, commerce and corporate security, our User Workshops have shaped the flexibility and operability required of the model, and have driven the requirements for clear and intuitive visualisation. In addition, through a large-scale participant recruitment exercise, our Participant Cohort will inform us regarding the social acceptability of our model, so we operate with awareness of ethical and legal frameworks.

Year Two is about consolidation of the SuperIdentity approach, through further empirical work and through optimal use of our framework to use known information either to predict unknown information, to link previously isolated indicators, or to direct intelligence-gathering. A large part of this work will be informed by our Home Office backed investment into data collection, yielding a considerable set of biographic, biometric and cybermetric measures from a single group of volunteers. Data management and ethical clearances are in hand to create a database, licensed to researchers, and capable of establishing constellations of measures that co-occur and combine to indicate identity. The novelty of the SuperIdentity project remains in the combination of measures across online and offline contexts so that identity in one domain can be cross-referenced with identity in the other.

Following from successful dissemination activities both to scientific communities, public audiences, and to stakeholders, Year Two sees our preparations for a Research Roadshow incorporating video material and hands-on demonstrations. These complement a series of user workshops, conference attendances, and scientific papers for academic dissemination.

3. The SuperIdentity Project

Our Context

In modern society, the risk associated with unreliable means of identification is felt in terms of a threat to personal privacy, information, intelligence, and resource. In the context of identity fraud, a recent assessment by the National Fraud Authority estimates the costs of UK identity fraud to exceed £2.7 billion per year, affecting 1.8 million people with much of this impact hitting the public purse. Allied to this, a government review commissioned in 2010 suggested that the capacity to obtain counterfeit identification documents contributed to the illegal entry into the UK of between 863,000 and 1.1million individuals, with a significant cost to national infrastructure and a potential threat to national security. Finally, failure to assure identification carries a cost in terms of criminal proceedings, and implications for the entire justice system. Indeed, identification of the wrong suspect can contribute to the criminal trial, conviction and sentencing of an innocent party, together with a failure to pursue the true perpetrator. Technological enhancement means that identity can now be revealed, and counterfeited, in complex ways both in the real and the cyber world in a manner that existing models of identity and identification cannot keep up with. The SuperIdentity (SID) project represents an urgent and necessary response to this issue.

Our Aims

SID offers an innovative and exciting new approach to the concept of identity. The assumption underlying our hypothesis is that whilst there may be many dimensions to an identity – some more stable than others - all should ultimately reference back to a single core identity or a 'SuperIdentity'. The obvious consequence is that identification is improved by the combination of measures. SID takes this approach further than any existing work, and we achieve this by including static and behavioural measures from both the real world and the cyber world. Indeed, as perhaps the fastest growing identity domain, and the fastest changing means of self-representation, cyber-identity must not be ignored in models of identity. In this way, SID will address not only the short and mid-term issues but also longer-term future developments in identity and identification.

In addition, SID provides two capabilities that are unique. First, we offer an identity framework through which associations can be made between different identity measures. The value of these associations is that one known piece of information may then be used to predict another previously unknown piece of information. This sort of approach is commonly used within e-commerce to enable analysts to predict that a shopper who purchased Product X might also be interested in Product Y. However, this approach has not been used previously in the realm of identity, and offers significant value to security and intelligence services.

Second, we offer the capacity to quantify the certainty associated with an identification decision. This enables the end-user to have a level of confidence (or risk) in their decision, and to make a judgement as to whether additional information is required.

Our Objectives

This work is of theoretical importance in and of itself. However, more than this, it will inform critical applied questions pertaining to the reliability and robustness of identification within modern and future societies. Our aims are expressed through three objectives:

(i) to combine identity measures across real and cyber domains to inform identification decisions in the face of partial and changing knowledge and uncertainty;

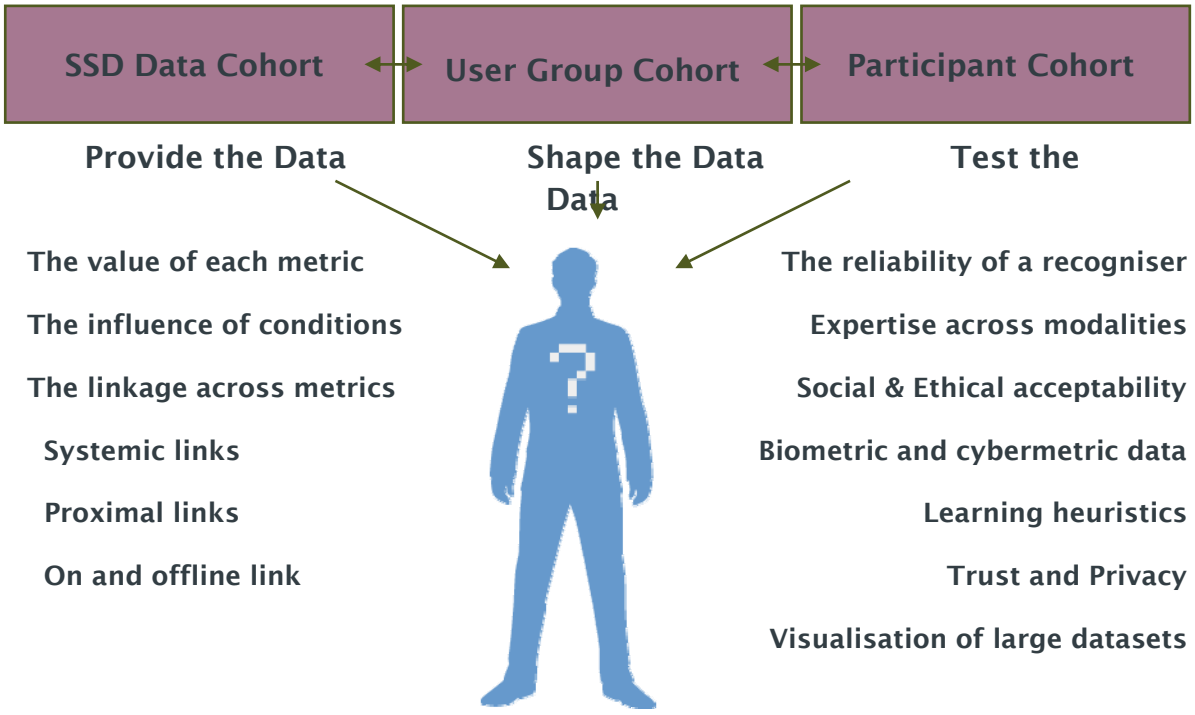
(ii) to uncover hidden data and relationships between data which can contribute to informed decisions about identity; and

(iii) to quantify the certainty of an identification by quantifying the reliability of each contributing measure.

4. The SuperIdentity Framework and Methodology

Our Methodology

The SuperIdentity methodology relies on the development of three important sources of data. These reflect the importance of metrics or cues to identity, the importance of user requirements when working with identity, and the importance of social understanding and acceptability together with legal awareness when enquiring about identity. These three sources of data are reflected in our overarching methodology for our Year One activities, and each is discussed below:



Southampton Stimulus Database (SSD)

The Southampton Stimulus Database is the culmination of 9 months of planning, data management assurance, and ethical clearance. It will deliver a database from 100 individuals who will provide biographic, biometric, cybermetric, and personality data. With all data points coming from the same individuals, the capacity exists to cross-correlate one piece of information with another, using meaningful and theory-driven expectations. The result will be the capacity to determine whether knowledge of one piece of information can predict other (related) pieces of information. Critically, this will enable a linkage of biographic, biometric and cybermetric indices of identity. Our suspicion is that various aspects of personality may underpin a critical linkage function. As an example, the personality characteristic of 'extroversion' may predict the biometric of long stride length, the cybermetric of a firm pressure in a mobile phone swipe gesture, or the cyberbehaviour of a large and active friend set online. If such a linkage is proven, then demonstration of any one metric can lead to (i) the prediction of the other

related metrics, (ii) the targeting of information searches, or the triangulation of identity across diverse cues.

Our Southampton Stimulus Data collection will commence in November 2012 with the support of the Home Office Centre for Applied Science and Technology (CAST), and with awareness of the Interpol and Custody suite standards for existing biometric data collection. As a key output, we will work with the Home Office on the preparation of a Standard for data collection. In addition, we are in the preparatory stages of a report to determine the reliability of our ten court biometrics (those that meet an evidentiary standard) so that best use of existing and novel biometrics can be made within the court system.

User Cohort

Our user-cohort has been recruited from amongst a group of professionals who regularly have need to identify individuals, or gather evidence, as part of their day-to-day roles. In the US, these individuals span the fields of Law Enforcement, Intelligence analysis, Border Control, Consumer Research, Fraud, and Corporate Security. In the UK, these individuals span various government agencies and commercial companies. None are named here.

The purpose of the User Cohort is to direct the functional requirements of the eventual SuperIdentity framework. Through semi-structured interviews, their insights into the desired capability of a SuperIdentity system have helped to inform both the data to be gathered in the Southampton Stimulus Database; and the flexibility and customisation of the SuperIdentity model itself. The outcomes of stage 1 of interviewing are summarised later in this document.

Participant Cohort

The participant cohort represents a special group of participants who will follow our project across a two year period. They will be recruited to take part in experimental studies, and this will enable us to address the possibility of 'super-recognisers' who are notable at recognising individuals within and across metrics.

However, the real value of our participant cohort is in their role in providing a social reflection on the acceptability of a SuperIdentity framework, and the levels of education or risk-taking that individuals show to their (super)identity information. The team at Bath have specific expertise in working with participant cohorts, and they bring this to bear in the recruitment, engagement, and involvement of a group of 15-18 year olds. This cohort represents an under-researched group of nevertheless high-traffic online users. Consequently, such a cohort provides the team with a very rich opportunity to learn about

the ethical and social acceptability issues concerned within a modern identity context. We are now at the stage of ethical clearance and cohort recruitment.

Legal Input

The SuperIdentity project has benefited from the input and insight of our legal experts. The Southampton Law team have provided a thorough appraisal of the state of the art in Identity research, and this has been presented alongside the very recent changes in legal reforms within the EU. Issues such as 'the right to be anonymous', the ownership of online data, and the Information Commissioner concerns of function creep (where information is obtained for one purpose but used also for another purpose), have informed the sorts of experimentation we have selected, and the sorts of questions that our participant cohort will be directed to consider.

In addition, the involvement of our legal team has ensured that we act responsibly within the SuperIdentity team in terms of the ethical clearances we obtain, and particularly in terms of our data management policies to safeguard the information within our Southampton Stimulus Database.

5a. Current Findings: User Perspectives

Engagement of our User Cohort has been championed by teams at PNNL, Bath and Oxford, with the assistance of our Steering Group members. The priority with this piece of work is to inform the SuperIdentity model so that we understand and prioritise the most relevant metrics or measures, and the most likely links between measures to support identification. The user perspectives thus provide a deeper context for our work, point to areas for additional work, and inform our requirement to create a valid user experience and a clear front-facing visualisation of identity and identification.

Users have been drawn from a broad US and UK community, spanning law enforcement, intelligence, border enforcement, fraud, corporate security and consumer research. Through semi-structured interviews, designed to be unclassified, broad themes emerged. First, it was noted by our users that not everyone under their scrutiny is suspected of being a 'bad guy'. In this sense, identification should be seen as one of a larger set of goals that they may possess. Second, the issue of online deception was flagged as important. The capacity to build confidence in identification is seen as desirable, through building constellations of corroborating evidence. Third, the issue of provenance is noted – a sense of knowing the source and thus the likely accuracy or reliability of information.

Each user also identified a series of other more tailored priorities that provide valuable context for the SuperIdentity project. For instance, in an intelligence or law enforcement context (i.e., investigation of foreign interests, investigation of immediate threat), there is a particular need for real-time information, whereas in a more corporate setting (i.e., investigation to confirm and profile company involvement) the onus is on consistency of information across sources rather than on the speed of obtaining that information. At the level of intelligence gathering, priorities may lie in profiling an individual and identifying real names, known associates, or potential affiliations where information may be sparse or deceptive, and this will have greater or lesser urgency depending on the reason of interest (i.e., cyber-attack). Equally, the intelligence arena has a need to determine the reliability of source information so that the provenance of any intelligence can be verified.

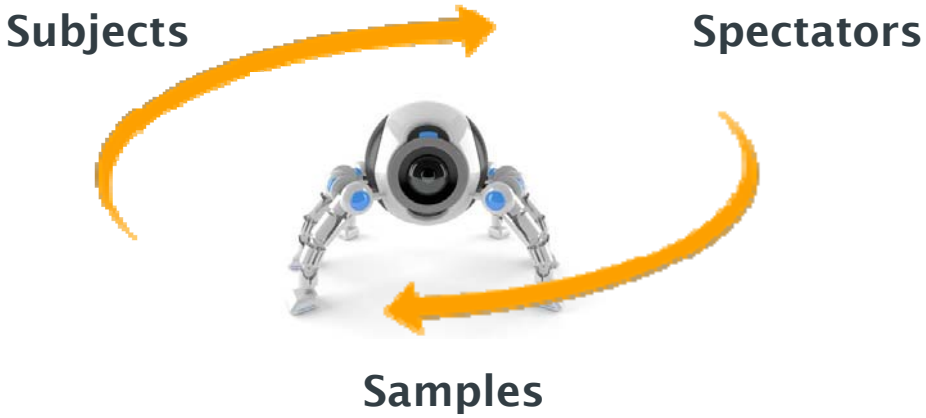
Together, these insights into the needs of the user and the requirements of the SuperIdentity framework, inform several aspects of our current work. They shape the metrics identified by users as being of high value, and thus included in our Southampton Stimulus Database. They articulate the needs of the user in linking fact A with fact B so support user goals, and thus feed into both hypothesized correlations within our model and within our current empirical study. They also underline the value of a combinatorial approach in which known information can be used to reveal hidden information, and can support the direction of additional intelligence gathering. All elements of our user perspectives are being captured in our on-going SuperIdentity work.

5b. Current Findings: Biometrics





Our approach to verify identity and identifications from real-world biometrics has recognised the importance of three possible sources of information:



The Spectator (who is doing the recognising ?)

The Subject (who is being recognised ?)

The Sample (what are they being recognised from ?)

Exploration of the Spectator is being conducted by the Southampton team. It will tell us whether the source of an identification is important. Should we, for example, trust an identification from source X more than an identification from source Y.

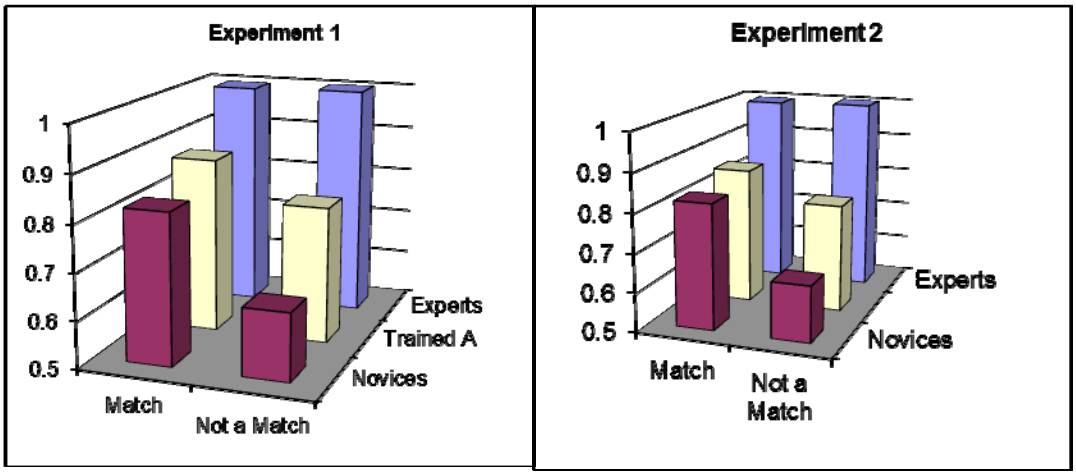
Exploration of the Subject is being led by the Dundee team. It will tell us whether some individuals might be more or less easily identified from a given biometric than others. Bound into this enquiry is the extent to which a given biometric has the capacity to reveal identity despite changes in an individual's age, health, genetics, occupation, or cumulative environmental influences.

Finally, exploration of the Sample is being provided by the Kent team. It will inform us about the importance of the sample quality or contextual factors that might make a given biometric more or less reliable. Automated techniques can quantify the extent to which identification may be compromised by a change in sample quality. Again, our purpose is to determine whether we can still trust an identification decision. Significant progress has been made in all three areas.

The Spectator:

Issue 1: Human Expertise

We have explored the field of human recognition as a way of determining whether the source of an identification (the spectator) makes a difference to the quality of that identification. In a series of studies, we used Fingerprint Experts as a target community. Our research suggested that fingerprint experts use an established methodology which is clear enough to explain to a novice, and can be used as a training tool. Importantly, this training tool enabled significant improvement in the capacity to scrutinise matching and non-matching fingerprint pairs, and elevated performance above the level of an untrained novice.



Reassuringly, the number of years of experience held by an expert had no effect on their competence – all experts were as good as one another regardless of personality type or years in the job. However, the experts were still significantly better than our trained novices and this may reveal the importance of the 'reality of an ecologically valid situation'. The experts took time and care to ensure their decisions were right because they were very aware of the consequences of an error in the context of their day-to-day job within the legal system. Our trained novices lacked this realism and their performance was not as good as the experts as a consequence.

Issue 2: Human Awareness

Our use of cognitive psychological techniques enabled us to explore not only how well an individual performed on a recognition task, but how well they believed that they performed. This becomes important in the absence of ground truth. In such a situation, how do we know if an identification is right ?

We recorded individual's self-evaluations through three measures : self-rated confidence, recollective experience, and willingness to 'report to the authorities'. The first of these is often used, and is simple to understand, but is fraught with interpretative problems because of differences in how confident people are *per se*, and in how they use the scale to record

changes in their confidence. The other measures were therefore of interest to us.

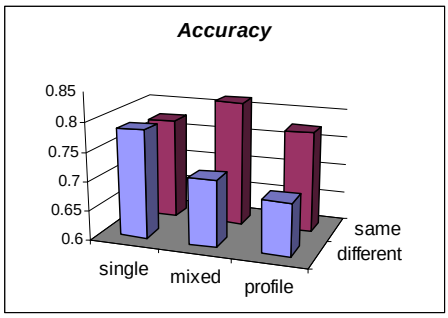
Our work showed that individuals monitored their performance well. As performance responded to the impact of facial distinctiveness, so did their confidence, recollective experience and reporting. The last measure showed the clearest effects though and our recommendation to users underlines the value of a report to an authority. Coerced identifications are less likely to be correct in this regard.

Issue 3 : Human Limits

We have also explored ways to identify the limits of recognition and here, we have used two different biometrics – the face and the voice. These provide quite different challenges and display quite different vulnerabilities to recognition error. The face is generally recognised very well whereas the voice is a relatively poor means of identification.

In terms of voice recognition, our results showed how easy it was to interfere with recognition ability, merely by inserting distractor voices in between target and test. It did not matter how similar or dissimilar those distractor voices were – performance fell significantly and immediately.

Face recognition was, however, capable of being enhanced. In a controlled study where participants were asked to see whether an approaching walker matched the person shown in a passport, our results suggest that performance was significantly improved when the passport incorporated a ¾ profile as well as a full-face image. This mixed image passport provided richer information from which to make the identification.



Full face Mixed ¾ Profile

Taken together, these findings suggest that some spectators ARE better equipped to recognise people given their domain expertise, but that spectators generally are aware of when they are right and when they are wrong.

The Subject

The question remains though as to whether our biometric measures are good enough and the conditions are sufficient to make an identification. Of particular importance is the

concept of whether a biometric measure taken at time 1 is still reliable as an indicator of identity at time 2. An understanding of the critical reliability of a measure will combine with our understanding of the reliability of the spectator, and together, both will establish the level of confidence in the final identification.

As part of this enquiry, the team at Dundee, Southampton and Kent have identified three sets of biometrics currently known for their evidentiary value within the perspective of English Law. These include (i) an established set of biometrics (i.e., DNA, fingerprints), (ii) an emergent set of biometrics (i.e., ear prints, iris scans, retinal scans and gait), and (iii) novel biometrics (i.e., hand vein patterns, hand geometry). To meet an evidentiary standard in court, the judge acts as gatekeeper to assess probative value and determine admissibility. However, each time a biometric is used in court, its reputational value is under scrutiny. Mistakes, in terms of the degree to which a biometric meets scientific standards and withstands cross examination, can damage its value for all cases that follow. These serve as a warning for emergent and novel biometrics and underline the importance of clear scientific thresholds and resilience under cross-examination. Our review of case lore and of scientific rigour will guide both the legal forum and the more investigative forum as existing biometrics are evaluated and new biometrics are explored.

The contribution of the Dundee team has also enabled another vital contribution to the work of the SuperIdentity project. Rather than consider each biometric in isolation, our understanding has been developed through an appreciation of the systemic linkages between biometrics (one biometric is linked to another through an underlying common biological system), and the proximal linkages between biometrics (one biometric is linked to another through an underlying physical proximity). This understanding enables the team to articulate not only the individual contribution of each biometric, given its reliability over time (see above), but also to articulate the linkage across biometrics so that knowing one piece of information can enable us to predict another. This combination of biometric information sits at the heart of the SuperIdentity approach, and has been captured through explicit prediction and through formal linkage in our SuperIdentity model.

Finally, the Dundee team have contributed vision and cutting edge development to our set of considered biometrics. Positioned at the forefront of their field, their appreciation of novel biometrics such as hand geometry and vein patterns means that these emergent biometrics gain visibility through peer review in the scientific community and through their place as evidence within the court process. The value for the SuperIdentity project is that individuals with malign intent may know to mask or disguise the commonly understood biometrics but may not yet be aware of the potential to reveal identity through more novel biometrics. The Dundee team put our awareness ahead of that of the uninformed criminal in this regard.

The Sample

The team at Kent have spearheaded our exploration of biometric identification by automated systems. This represents a perfect complement to the human identification work and indeed, our experimentation has been run in parallel to explore not only the same biometrics, but actually the same individuals from these biometrics, and the same contextual conditions of viewing. The team at Kent have delivered an understanding of biometric identification using publicly available algorithms.

Whilst some variability across algorithms is evident, the results reveal a capacity to recognise individuals from a facial image despite change in the noise of the image (i.e., background details, others visible) and changes in camera, lighting and viewing distance. These factors have a predictable and measurable effect, and suggest that optimal automated face recognition will be obtained when viewing distance is minimised (as long as the camera placement is such that the full face (rather than a top of head) view is provided). When matching a face from an unconstrained photo to a stored full-face referent, performance was best when the angle of rotation was small. Rotation beyond about 22.5 degrees caused a notable and significant decline in both extraction and verification of the face within the dataset. Finally, automated recognition was significantly affected by the size of the watch list being searched. The smaller the watch list, the better the performance and, of course, as the watch list increased in size, facial angle, distance and camera variation then had minimal effect as performance declined towards chance levels.

These results allow us to quantify the likely accuracy of automated identification from a facial image, and the likely impact of known factors which compromise identification. These meta-data are reflected in the SuperIdentity model as 'context factors' which make identification from a given biometric more or less reliable. Here, this change in reliability can be measured and indexed.

What is striking, however, is that even in the best of viewing conditions, and with the best of the tested algorithms, automated facial identification only approached 30%. As a response, the Kent team advocate exploration of a combination of algorithms or engines so that identification decisions can benefit from triangulation. Even more powerful will be the integration of human heuristics for face recognition into the development of smart automated systems, and work is underway across Southampton and Kent teams to explore this avenue.

5c. Current Findings: Cybermetrics





Offline and Online Identities

Several of our teams have been involved in the exploration of identity across offline and online contexts. The question here has been 'how do people represent themselves in different settings?'

We have used the Twenty Statements Test to probe this question. It allows individuals to describe themselves in twenty statements, and we then ask whether they are happy to reveal their answers or whether they want to withhold or replace anything they have said. Our results demonstrate that people represent themselves very similarly across an offline and an anonymous online context. The latter may provide a sense of safety so that, despite having information visible in an online setting, individuals do not know who is looking at it so they feel no judgement or scrutiny, and thus no need to regulate their image.

In contrast, when individuals represent themselves intentionally in specific online spaces, such as a dating site, or a professional site, then they tend to express aspects of their self that may be 'ideal' for that context.

The pattern of information withheld was also interesting in as much as this may indicate what people would lie about if they had no capacity to withhold. Individuals tended to withhold subjective judgements about themselves rather than the more easily verifiable facts. Surprisingly though, participants who were more socially self-aware did not withhold more information. In contrast, perhaps individuals in an online space were not affected by the traits that affected them in an offline and less anonymous context.

Together, these data suggest subtle differences in how identity is managed in offline and online contexts: Changes in the rules of engagement across these contexts fundamentally affects rules of impression management and display, and we should not assume that offline and online contexts are equivalent.

Smart Phone Gestures

To a large extent, the gesture-driven touch-sensitive interactive screen has removed the need for physical buttons to interact with mobile phones. As highly sensitive instruments, touchscreens are able to provide researchers with access to more nuanced data about user interactions than could be obtained from two-state physical buttons and keypads.

Ongoing work by the team at Bath has explored the use of multiple 'swipe' gestures for the purposes of identification. Gestures were captured during user-interactions in four directions from a wide range of mobile smartphone users. Using four simple feature extractions *gesture length, completion time, touch pressure* and *gesture thickness* we were able to distinguish users by their gender, age range and by the hand used to create the swipes. By using cluster analysis techniques, we were further able to classify swipes into three distinguishable 'styles', based on contributions from all four feature extractions described. Finally, by examining how consistently each user created swipes within these styles, we found that all of our participants naturally created their swipes using no more than two of these styles. These findings are explored in terms of their potential utility for passive user verification and user identification via swipe gesture characteristics.

Password Riskiness

Colleagues from the University of Leicester and University of Oxford are conducting research that aims to a) investigate the ways in which individuals create and keep passwords, and b) explore the individual differences in password-use in relation to personality and internet experience. We are currently awaiting approval from the ethics committees at both institutions and hope to start data collection by the end of September 2012. The online questionnaire-based research involves observations of actual password creation together with self-reported behaviour relating to the way individuals create and manage passwords across a variety of different services, both in an online (e.g. email, Facebook) and offline (e.g. mobile phone and credit card pins) domain. The research is split into a series of studies with an aim of refining a number of hypotheses. We make a number of predictions including: aspects of personality will affect the strength of passwords created by participants; those that use the internet regularly will create secure passwords (passwords that are harder to crack), but will use them in multiple domains; password strength and management of that password will be dependent on the data they are protecting; and secure passwords will be less likely to be remembered. This research will add to the growing body of work on password behaviour, and allow the SID project to make confident associations between aspects of personality and password behaviour.

Social Education: Feeding back on online security choices

Drawing upon the work completed on the Twenty Statements Task, the team at Bath are currently piloting a new study that again examines the degree to which people elect to reveal personally identifiable information about themselves in their online personal profiles. Following the design of the Twenty Statements studies, participants are required to create a

personal profile based on either an online dating or professional networking scenario.

Where this study differs from previous work is in its use of the emerging SuperIdentity model as an experimental device. Upon submission of their profiles, known SuperIdentity elements (e.g. email address, family name) are extracted by the researchers and their potential SuperIdentity linkages (e.g. company name, position) are determined via an API link to the overall model. A breakdown of these links, and the information used to derive them, is then returned to the participant as a warning notice, and they are given the opportunity to revise their submission based upon this feedback.

The research questions of this study are twofold: 1) How do people choose to manage what they say about themselves online when they are presented with their feedback? Do they choose to omit that information entirely, reduce its visibility (e.g. 'show only to friends') or do nothing at all? 2) To what degree is this behaviour affected by the scenario within which the profile is created? For example, are people more likely to self-censor within an online dating website than in a professional networking website?

Participant recruitment for this study is expected to begin in early October 2012.

Deception Study

The team at Leicester are also conducting a research which examines differences in deception in an online and offline context. This study uses longitudinal self-report methodology. Over the course of a week, participants provide information each day about the lies they have told, whether these be ‘little white lies’, or lies that they consider to be ‘more serious’ in nature. Participants also complete personality measures.

Ethical approval is in hand, and recruitment is expected to begin in October 2012. This project will allow the SID project to make predictions as the veracity of online self-disclosure, and this gains importance particularly in the case where online and offline information may be at odds with one another, or may be at odds with intelligence gathered through an alternate route.

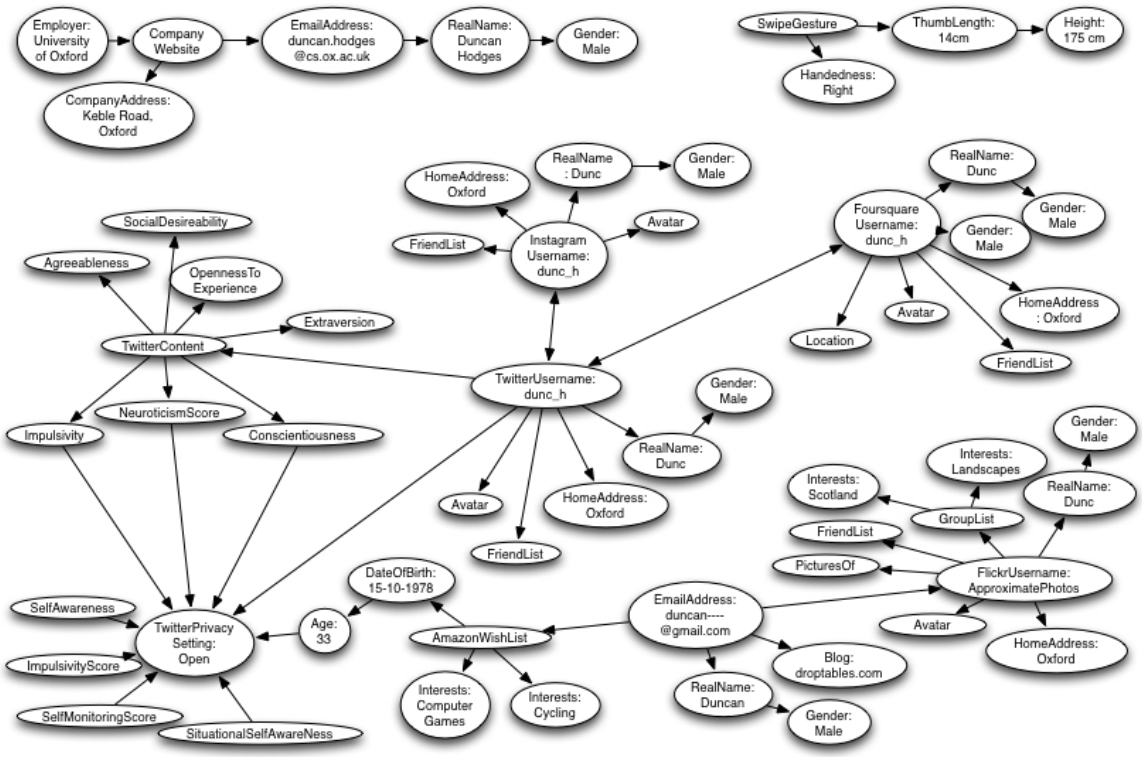
6. Fusion and Visualisation

The Combinatorial Model

The Oxford team have been instrumental in creating a mathematical model, based on Bayesian principles, which allows information to be combined so that logical questions can be asked. For example, given facts A and B, can I find out C?; and given a desire to find out fact C, what information do I need ? This model enables the SuperIdentity team to fulfil its brief in weighting the value of information, the source of information, or the contextual influences on information. This enables an index of certainty to be attached to an identification decision. The model offers the intelligent capability to go further. Specifically, we are able to use known information to predict previously unknown information, and we are able to direct information-gathering to support our identification decisions. Both have huge intelligence benefits.

Towards the end of Year One, significant energies have been devoted to the development of this SuperIdentity model. It possess a hierarchical organisation in which elements may be defined which nevertheless can be recognised through their sub-components. For example, a face may be recognised via its sub-components of ear shape, iris colour, distinguishing marks, hair colour and style, etc. The model is configured to enable elements and their sub-components to be linked according to systemic or proximal commonalities, or according to behavioural consistencies.

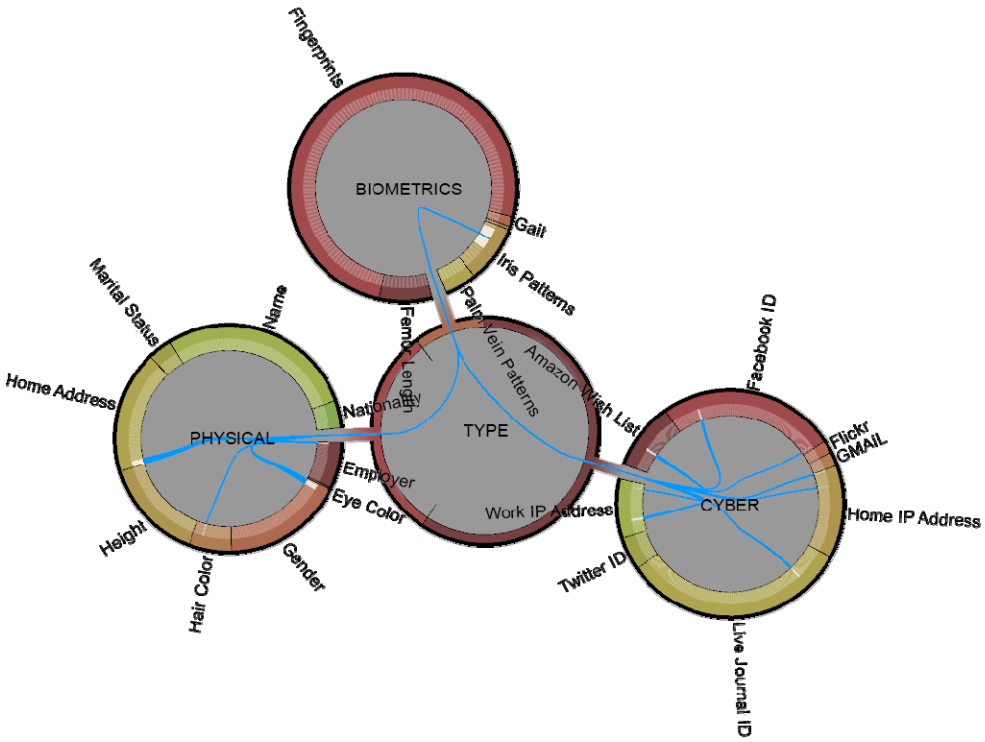
Key in this model is the value of individual personality characteristics because these, we believe, play a significant role in shaping dynamic identity cues both in the offline world and the cyberworld. These, consequently, become an important and perhaps unspoken route through which biometrics may predict cybermetrics and *vice versa*. The SuperIdentity model allows articulation of all these possibilities.



In addition, the SuperIdentity model is configured to reflect the strength of relationships, and the evidence base underpinning those relationships, be these through scientific findings or through expert heuristics. As a consequence, both the strength or confidence of our intelligence and the provenance of our information can be verified, satisfying the user needs as articulated in our semi-structured interviews.

The Visualisation

The PNNL team, in combination with HCI colleagues at Bath, and modelling colleagues at Oxford, have developed an exciting framework within which the SuperIdentity model can be visualised. This visualisation tool is again guided by both the needs of the user community, whilst respecting the power within the data of the SuperIdentity model itself.



The *Arcweld* visualization emphasizes the potential of the SuperIdentity Model. *Arcweld* is a radial visualization that accentuates the relationships that may exist between elements - even across the layers of hierarchy. By grouping elements first by their cyber, biometric and natural world designations, we can see the highly desirable transformations capable of crossing the chasms between these worlds. Digging deeper, we can discover all relationships to a particular element. While the *Arcweld* illustration (above) does *not* reflect actual data or transformations available in the current model, it does help to demonstrate the power of the model. As an example, imagine that "Iris Pattern" was known. The Iris Pattern element shows links to several elements in the Cyber domain and in the Physical domain. Consequently, given one's Iris Pattern, we can trace both a Home IP address and a home address in a physical sense. This demonstrates the power and utility of the *Arcweld* visualization tool which has its strength in the capacity to link elements (known and unknown) via a series of associations or transformations to enable identity attribution.

7. Dissemination

Online Activities

Website: www.superidentity.org (877 unique visitors)

Links to: IMPRINTS: http://www.imprintsfutures.org/links/

Outreach and Dissemination

Guest, R.M. (2012). The SuperIdentity Project: exploring relationships between physical and cyber identity domains. Biometrics Institute, New Zealand High Commission, London. Sept 13th 2012.

Hodges, D. (2012). Geek Night, University of Oxford.

Stevenage, S.V., & Neil G.J. (2012). Representing yourself online. Interactive stand and dissemination materials at Community Open Evening: INTECH Science Centre, Winchester, Hampshire

Stevenage, S.V., (2012). CSI day for Year 8 students. How can you tell who someone is? Delivered to 76 Gifted and Talented local school children under the Southampton Learn with US outreach programme.

Academic Conferences

Bevan, C., & Stanton Fraser, D. (submitted). Touchscreen Biometrics: What Do Your Touch Gestures Say About You.

Hodges, Duncan; Creese, Sadie; Goldsmith, Michael. (2012) "A Model for Identity in the Cyber and Natural Universes," *Intelligence and Security Informatics Conference (EISIC), 2012 European* , vol., no., pp.115-122, 22-24 Aug. 2012 doi: 10.1109/EISIC.2012.43
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6298821&isnumber=6298809

Emanuel, L., & Jamison-Powell, S. (submitted). Looking after Me, Myself and CHI: Identity and Cybersecurity.

Saxby S. (2012). The SuperIdentity Framework. *7th International Conference on Legal, Security and Privacy Issues in IT Law* (LSPI), 2-4 October, Athens, Greece.

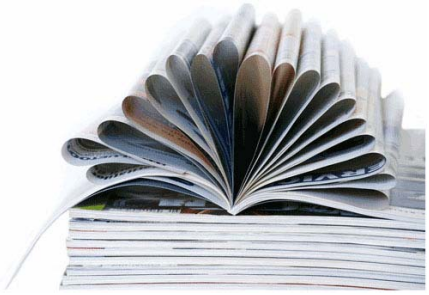
Saxby S. (2013). SuperIdentity workshop. *8th International Conference on Legal Security and*

Privacy Issues in IT Law (LSPI).

Stevenage, S.V., & Neil, G.J. (2012). Knowing What you Know: Using Metamemory to Predict Accuracy of Eyewitness Identifications. *IA-IP*. 5-7 December, London.

Stevenage S.V., & Neil, G.J. (2012). The relative strength of voices and faces in person recognition. *British Psychological Society, Cognitive Section Annual Conference*. Invited talk within the Voice Recognition Symposium. 29-31 August, Glasgow.

Academic Publications



Bevan, C., & Stanton Fraser, D. (submitted). Touchscreen Biometrics: What Do Your Touch Gestures Say About You.

Black, S.M., Creese, S., Guest, R.M., Pike, B., Saxby, S., Stanton Fraser, D., Stevenage, S.V., & Whitty, M.T. (2012). SuperIdentity: Fusion of Identity across Real and Cyber Domains. Refereed Proceedings of *ID360: Global Issues in Identity*, Texas, April 23-24th, 2012.

Hodges, Duncan; Creese, Sadie; Goldsmith, Michael (2012) "A Model for Identity in the Cyber and Natural Universes," *Intelligence and Security Informatics Conference (EISIC), 2012 European* , vol., no., pp.115-122, 22-24 Aug. 2012 doi: 10.1109/EISIC.2012.43 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6298821&isnumber=6298809

Stevenage, S.V., Neil, G.J. et al. (2012). The effect of distraction on face and voice recognition. *Psychological Research*, DOI: 10.1007/s00426-012-0450-z

Stevenage, S.V., Neil, G.J. et al. (in press). Recognition by Association: Within- and Cross-modality Associative Priming with Faces and Voices. *British Journal of Psychology*

Press Coverage

EPSRC: New Grants Announcement: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/J004995/1

Southampton: Home Page story: University of Southampton, (17,148 hits) http://www.southampton.ac.uk/promotion/super_identity_01.shtml

Press Release: 18th January 2012 http://www.southampton.ac.uk/mediacentre/news/2012/jan/12_07.shtml

Oxford Press Release: http://www.cs.ox.ac.uk/projects/superidentity/

News: http://www.cybersecurity.ox.ac.uk/projects.html

CS centre: http://www.cybersecurity.ox.ac.uk/projects.html

Bath News: http://www.bath.ac.uk/psychology/research/castl/create-lab/

Press Release: http://www.bath.ac.uk/news/2012/01/19/superid/

Leicester Press Release: January 2012: Leicester University

http://www.southampton.ac.uk/mediacentre/news/2012/jan/12_07.shtml

http://www2.le.ac.uk/offices/press/press-releases/2012/january/a-web-of-lies

News Story:

http://www2.le.ac.uk/departments/media/research/research-groups/digital-identities-research-group

Dundee News Story: http://www.lifesci.dundee.ac.uk/cahid/human-identification

Press Release: http://www.dundee.ac.uk/pressreleases/2012/january12/reducecrime.htm

News: http://www.biodundee.co.uk/index.asp?tm=3&nid=733

News: http://www.southampton.ac.uk/superidentity/who/biographies/sueblack.page

Kent Press Release: http://www.kent.ac.uk/news/stories/kent-contributes-biometrics-expertise-to-super-identity-model/2011

News Story, p8: http://issuu.com/universityofkent/docs/kent1112

Loughb'gh http://www.inloughborough.com/news/100646/A%20web%20of%20lies

The Deadman: http://www.thedeadman.co.uk/category/blog/

Gemini: http://ip-192com.blogspot.co.uk/2012/01/super-identity-project-focuses-on-real.html

StumbleUpon: http://www.stumbleupon.com/content/1JlQuC/likes

ip-192: http://www.ip-192.com/2012/01/18/cyber-identities/

Motherboard US: http://motherboard.vice.com/2012/1/31/you-are-what-you-post-hunting-the-elusive-superidentity-q-a

PearlTrees: http://www.pearltrees.com/#/N-f=1_3823659&N-fa=3256918&N-p=30055166&N-play=0&N-s=1_3823659&N-u=1_357281

New Boundaries, University of Southampton: Trail piece Nov 2011; Full story Nov 2012