The Language of Biometrics: Analysing Public Perceptions

Oliver Buckley[a], Jason R. C. Nurse[b]

[a]School of Computing Science, University of East Anglia, Norwich, NR4 7TJ, UK
[b]School of Computing, University of Kent, Canterbury, CT2 7NF, UK

Abstract

There is an increasing shift in technology towards biometric solutions, but one of the biggest barriers to widespread use is the acceptance by the users. In this paper we investigate the understanding, awareness and acceptance of biometrics by the general public. The primary research method was a survey, which had 282 respondents, designed to gauge public opinion around biometrics. Additionally, qualitative data was captured in the form of the participants' definition of the term *biometrics*. We applied thematic analysis as well as an automated Word Vector analysis to this data to provide a deeper insight into the perceptions and understanding of the term. Our results demonstrate that while there is generally a reasonable level of understanding of what biometrics are, this is typically limited to the techniques that are most familiar to participants (e.g., fingerprints or facial recognition). Most notably individuals' awareness overlooks emerging areas such as behavioural biometrics (e.g., gait). This was also apparent when we compared participants' views to definitions provided by official, published sources (e.g., ISO, NIST, OED, DHS). Overall, this article provides unique insight into the perceptions and understanding of biometrics as well as areas where users may lack knowledge on biometric applications.

Keywords: Biometrics, thematic analysis, word vector analysis, security, privacy, perceptions, usable security, user study

1. Introduction

Biometric technologies are becoming increasingly commonplace in our everyday lives across a wide range of applications from securing our personal devices through to managing physical access. For example, a study by Techpinions [1] revealed that Apple's Touch ID fingerprint technology is used by 89% of users, with a Touch ID capable device. More recent research by Deloitte discovered that 79% of UK smartphone owners in general (i.e., iPhone and others) use their device's fingerprint scanner, and more than a third of smartphones now have a fingerprint reader [2].

It is entirely possible that users are now being exposed to biometric technologies without ever realising it. Since 2016, Barclays Bank has used voice recognition software for its personal telephone banking customers, and other banks such as HSBC have also introduced the technology [3] [4].

As our reliance on the Internet increases as does the requirement to create memorable, yet hard to guess security credentials. Research conducted by Grawmeyer and Johnson [5] in 2011 found that an average user will have to manage 7.95 passwords. In contrast to this, a survey carried out in 2014 as part of Cyber Streetwise (now known as Cyber Aware [6]), a UK government initiative designed to drive behavioural change in cyber security, found that an average user will manage 19 different passwords. This highlights a significant growth in the number of credentials a user is required to manage, with the number of accounts more than doubling in three years. Additionally, there is also the added overhead of authenticating to a wide range of services. Wash et al. [7] found that on average a user will typically be confronted with a password event anywhere between 8 and 23 times per day.

According to Gehringer [8], it is notoriously difficult to create memorable yet secure passwords. A strong password can be effective but it is all too common for this protection to be compromised by the users themselves. For example, a study by SplahData [9] reveals that '123456' and 'password' are still amongst some of the most commonly used passwords. The more traditional methods of security will typically focus on something the user knows (e.g. a username and password) or on something the user possesses (e.g., an RSA token [10]). Conversely, the use of biometrics can potentially alleviate many of the issues surrounding security credentials, as the emphasis is placed on what the user is, rather than what the user knows or possesses. When compared to more traditional approaches to authentication, biometrics claim to offer a greater level of confidence that the individual is who their credentials claim them to be. Additionally, according to Jain et al. [11] biometrics can potentially prove to be more reliable than standard authentication methods and do not rely on credentials or tokens that can be inadvertently lost or stolen.

With our ever-increasing reliance on digital services and the ubiquity and affordability of biometric technologies then it would seem that the solution is obvious. The use of these biometric technologies should make it easy to deliver user authentication that is both secure and cost effective, while at the same time reducing the cognitive load on the users of these systems and services. A crucial part of any increased use of biometrics however, is their acceptance by the general public.

Despite Apple's impressive statistics on user uptake there are still a number of common misconceptions about the way

in which biometric authentication actually works, as discussed by both Ashbourn [12] and Thompson et al. [13]. For example, concerns about biometric data being lost or compromised are commonplace, with users often assuming that fingerprints are stored exactly within their phone as an image. However, the reality is that when collecting a fingerprint the data is encoded in some way based on a number of extracted features. These fundamental misconceptions about this technology, particularly when it is so widely available, suggest that despite the increased prevalence of biometric technologies that they are still not widely understood.

In this paper therefore, we engage with members of the general public to investigate their understanding, awareness and acceptance of biometrics. Our goal is to provide useful insight into these factors as well as key related technologies. The remainder of this article is structured as follows. In Section 2 we provide a discussion of the related material, including previous studies of user perceptions. Section 3 provides an overview of the methodology used to conduct the study. Section 4 presents the results and discusses their significance, and finally Section 5 concludes this research.

2. Related Work

Furnell and Evangelatos [14] present a survey that explored the awareness and perceptions of biometrics. The survey looked to understand which techniques were commonly understood, which had been used and how they could be used in practice. This work highlighted that the respondents (of which they were 209) were generally positive about the notion of biometrics but with relatively limited practical experience of using them. Additionally, they found that there were a number of technologies that were generally better accepted than others. For example, participants were general more comfortable with the use of fingerprints than retinal scans. While it is over a decade since this seminal work, it will provide an interesting comparison with the research presented in our current article. As previously discussed, there is a greater exposure to biometrics in our daily lives and as such it is anticipated that the acceptance of biometrics will have changed accordingly.

Work presented by Chan and Elliot [15] provides an updated look at the privacy perceptions of biometrics using two surveys. The first survey, with 200 participants, asked respondents about their experiences and perceptions of biometrics. A second survey, carried out a year later, looked to measure any changes in perceptions over the course of time. This research also suggests a level of skepticism around the security and privacy of their biometric data. For example, over 45% of respondents would not trust their data with a public corporation. One of the key findings in this research was that there was a greater support for the use of biometrics in both counter-terrorism and banking.

The work of Sabharwal [16] focuses explicitly on the perceptions of biometrics in banking customers. This research used a survey to understand the concerns, opinions and perceptions of banking customers with respect to e-banking. The results from the survey suggest a number of key metrics when considering the large-scale deployment of biometric technology within e-banking including: reliability, performance, resistance to circumvention and privacy issues.

One of the areas where there has been a significant increase in the ubiquity of biometrics is in our own personal devices, with fingerprint recognition being very much a standard feature on a modern smartphone or tablet. Bhagavatula et al. [17] conducted a lab study and survey to determine the usability of smartphone specific biometrics (e.g., fingerprint and facial recognition). Their work highlighted that the majority of users actually preferred the use of fingerprint unlock over facial recognition or the use of a PIN. Similarly, it was found that users perceived fingerprint unlocking to be more secure and convenient than the use of a PIN.

Research by Krol et al. [18] focuses specifically on the acceptability of face biometrics as a replacement for CAPTCHAs [19]. The work used a lab study to test a range of human verification mechanisms, and then used surveys and interviews to determine the acceptance and perceptions of the various techniques. One of the key findings of this work was that the users were generally concerned about the use of their own personal image in verification, highlighting that the privacy of personal data is a key concern to potential users of biometric identification.

Ogbanufe and Kim [20] focus on user perceptions of the differences between biometric and more traditional methods of authentication for e-payments. Their work found that the biometrics authentication method significantly influenced the security concerns of an individual as well as the perceived usefulness and trust of the online store. It is interesting to note that their research found that users considered fingerprints to be more secure than a combination of credit card and PIN.

The use of biometrics is not limited to a user's personal life and these are technologies that are becoming increasingly common within the workplace. Carpenter et al. [21] presents a study focusing on the privacy concerns of employees related to organisational use of biometrics. Their results highlighted that the self-construal (the extent to which the self is defined independently of others or interdependently with others [22]) played a significant role in the formulation of privacy, perceived accountability and perceived vulnerability concerns. Their work suggested that they also act as notable indicators of the user's attitude towards biometric technologies in the workplace.

One of the common themes that is notable across the studies is that users have concerns about the privacy and security of their own personal data. This is something that is further investigated in the research presented in this paper as it explores the concerns of the participants and the contextual nature of these concerns. For instance, this work investigates whether the situation determines the level of security that a user feels is required. The current research to date, in the related literature, has been mainly focused on the perceptions of usability, security and privacy of biometric technologies. The work presented here aims to draw out the public perceptions of biometric technologies and how these perceptions are linked to acceptance. Previous work in this area has not attempted to ascertain users understanding of the term *biometrics*. In the study presented here,

we have built upon a traditional survey with the use of thematic analysis [23] and a vector representation of words [24] to better understand the term *biometrics* and its implications for users.

3. Methodology

To structure our research, we adopted a methodology consisting of common data gathering and analysis processes. Prior to our study commencing it was also reviewed by our university's ethical review board. For recruitment, we used a mixture of convenience and snowball sampling [25] to gather participants from the general public. This involved advertisements on social media including Twitter, LinkedIn and Reddit. Surveys were selected to allow the gathering of data; the primary reason for this being their ease of deployment and larger participant reach. We designed our survey such that questions first covered a range of demographics including: Age, Gender, Highest level of education, and Area of work or study. On the topic of biometrics, we asked participants whether they had some understanding of the term, and if they did, they were asked to provide a definition.

The survey contained a further eight questions that focused on the participant's awareness and perceptions of various biometric technologies. We asked participants to rank five generic situations (or usage scenarios), chosen to provide different data contexts, based on their requirement for 'security'. Specifically, the situations were Banking, Online shopping, Airport, Mobile device, Home. The choice of these scenarios was arbitrary and primarily motivated by having a set of scenarios that could be ordered by their perceived need for security. We view a potential ordering of these scenarios as follows from most to least: Banking, Airport, Home, Mobile Device, online shopping. Participants were then queried about which biometrics they would be comfortable using in each of the five situations. This was used to develop an understanding of the technologies that the participants viewed as the most secure and those that they viewed as the least secure. The survey next focused on the perceptions of the security of biometrics when compared to other common techniques, such as passwords and two-factor authentication.

Our approach to data analysis involved a combination of quantitative and qualitative techniques. First we applied statistical methods to analyse responses to close-ended questions. Tests for correlation across some aspects (e.g., between education level and familiarity with various methods) were also conducted using the Pearson Chi-Squared test.

Next, the definitions of biometrics provided by each participant were analysed using both manual and automated qualitative techniques. Thematic analysis is a manual data analysis technique, which focuses on allowing the assessment of qualitative data for common themes and patterns [23]. This a well-known technique that has been applied across a variety of fields. In addition to this approach we also made use of an automated analysis method, to provide a comparison. We used an approach called Word2Vec [24]; this technique models words in a vector space, allowing for additional insight into textual content.

4. Findings and Discussion

4.1. Survey Results

There were a total of 282 participants in our study, which is in line with other similar studies discussed in Section 2. As part of our analysis we used the Pearson Chi-Squared test to establish if any correlations were present in the data, however, we found there were no significant correlations of note.

The majority of respondents to the survey were aged between 35 and 44 (40%), as shown in Figure 1, with approximately 83% of all participants being aged under 45. However, this distribution of the participants' ages means that the vast majority of respondents will have either grown up with technology from an early age or been early adopters of new technologies.

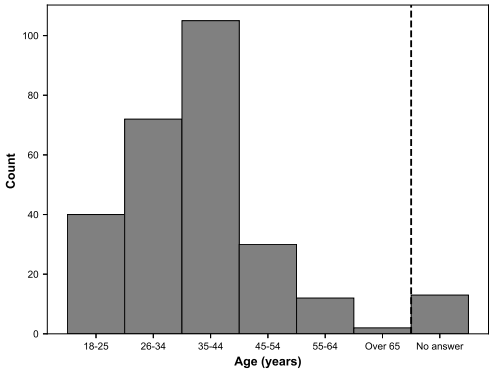

Figure 1: Age of respondents

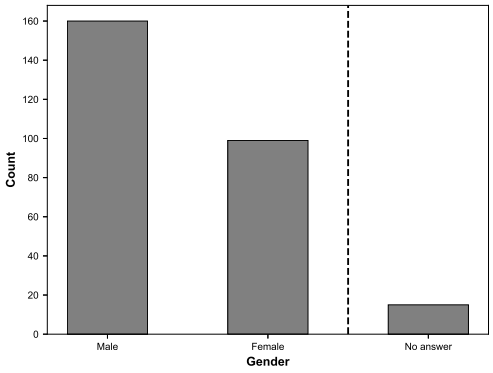

Figure 2: Gender of respondents

Figure 2 shows that most of the respondents were male (62%). In terms of education, a majority of participants were educated to at least degree level, with more than 67% having at least a Bachelor's degree and nearly 8% of all respondents held a doctorate, again this is likely to be influenced by the researchers' personal and professional networks. Our sample also represented a wide range of employment sectors. These include Accounting/Finance, Administration, Engineering, Human Resources, Education, Legal and Sales/Marketing. There

is a fairly even distribution among all of the sectors (ranging from 2-7%), with the exception of IT/Computing, which contains a clear majority (27%).

One of the first areas we examined was participant's knowledge of biometric systems. We presented participants with a list of common biometrics (both physical and behavioural) and asked them to indicate if they had heard of the scheme (Figure 3 and Table 1), whether they had knowingly used it (Figure 4) and whether they would be comfortable with it being held by a company or government (Figure 5).

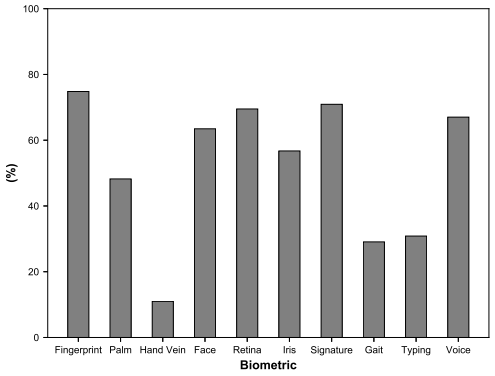

Figure 3: Have you heard of this biometric?

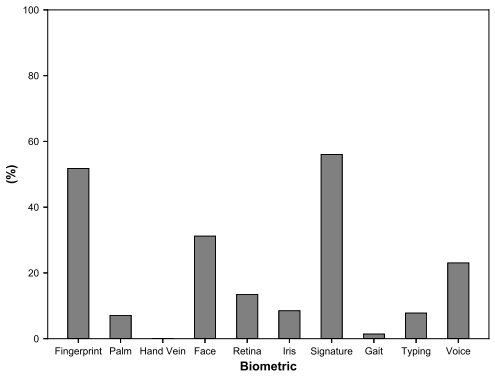

Figure 4: Have you ever knowingly used this biometric?

The first thing to note is that broadly speaking the participants had a good knowledge of a number of physical biometric technologies, as seen in Figure 3. The most commonly known of technology is the use of fingerprints, which is unsurprising given their prevalence in personal devices and our daily lives (e.g., smartphones or immigration at an airport).

At the other end of the scale, those that fewer participants had heard of, were gait, typing and hand vein. It is interesting to note that two of these methods (gait and typing) are arguably classified as behavioural biometrics, which perhaps suggests that there is a lack of awareness around these approaches. This is potentially unsurprising in that traditionally behavioural biometrics do not require the user to interact with any specific

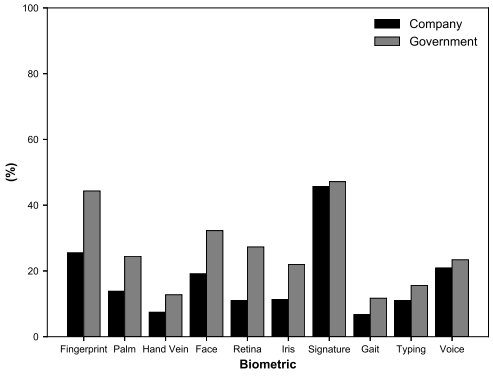

Figure 5: Would you be comfortable with a company or government holding this data?

hardware directly. Instead their behaviours are normally monitored remotely. This is further underlined by comparing the results to work of Furnell and Evangelatos [14], where keystroke dynamics also proved to be amongst the biometrics of which the fewest participants were aware.

The results in Figure 4 and Table 1 highlight the methods that have been knowingly used by participants. Some of the most commonly used are fingerprint and facial recognition. Both of these technologies have been used to secure personal devices and so such a common usage is expected. However, the most surprising result is that the most commonly used biometric was 'signature'. It is highly likely that many participants have mistaken a simple signature (e.g., signing for a parcel) with the biometric method. This is reinforced by Furnell and Evangelatos [14] who identified a similar case of potential confusion. Very few participants claimed to have knowingly used typing, voice and gait as biometrics. This is quite an intriguing point to draw out as these are indeed all biometrics that can be recorded remotely, so it is entirely possible (and likely in some cases) that these have been used by the participants without them actually having realised it. As an example, a number of banks are using voice recognition as part of their online banking services (e.g., Barclays Bank [3] and HSBC [4] in the UK).

Figure 5 (and Table 1) provides a comparison between the data that participants would be comfortable being stored by a government or a private company. From the results, we can see that respondents were uniformly more comfortable with their biometric data being held by a government, rather than a private company. This is a particularly interesting when considering fingerprints, which although potentially held on a variety of governmental databases, it is highly likely that a significant number of participants already trust their fingerprint data to a number of a device providers (e.g., Apple, Google or Samsung). This in itself raises other concerns, particularly in the future when we consider users understanding privacy implications of widespread adoption of biometric technologies [26].

The next question looked to understand which biometrics were considered to be the most secure. In the first instance participants were asked to rank five situations in order of their

			Comfortable held by:	
Biometric	**Heard of**	**Knowingly used**	**Government department**	**Private company**
Fingerprint	211	146	72	125
Palm print	136	20	39	69
Hand vein	31	0	21	36
Face	179	88	54	91
Retina	196	38	31	77
Iris	160	24	32	62
Signature	200	158	129	133
Gait	82	4	19	33
Typing	87	22	31	44
Voice	189	65	59	66

Table 1: The breakdown of which biometric methods a participant had heard of, knowingly used and would be comfortable being held by a government department or a private company.

	Banking	**Online Shopping**	**Airport**	**Mobile Device**	**Home**
1	85	2	81	16	18
2	90	28	42	21	21
3	25	69	36	39	33
4	2	54	24	83	39
5	0	49	19	43	91
Average Rank	**1.72**	**3.59**	**2.29**	**3.57**	**3.81**

Table 2: The breakdown of rankings for each situation. Each cell shows the total number of participants who had ranked the situation with that specific need for security (1 to 5). A total of 202 participants answered this question.

need for security, which allowed a ranking of the scenarios to be calculated drawing out those situations that were perceived to have the greatest need for security. An average ranking was calculated for each of the different situations; we note here that 202 out of the 282 participants chose to answer this question.

First the number of participants that had ranked each situation at any given level was calculated. This gave a total ranking for each of the situations, where the lower number dictates the situation with the perceived greatest need for security. This was then used to calculate an average ranking for each situation, and determine the situation that was deemed to have the greatest need for security and the situation with the least need for security. The breakdown of the resulting rankings of each situation can be seen in Table 2, with the overall rankings as follows: (1) Banking; (2) Airport; (3) Online shopping; (4) Mobile device; and (5) Home.

The results of this ranking are intriguing for several reasons. It is perhaps not surprising that banking was ranked as the situation with the greatest need for security. However, the fact that home was ranked as the least need for security was particularly surprising as our own home is the place where we are meant to feel the most safe and secure. Further investigation revealed that approximately 58% of respondents actually ranked their mobile device as requiring a greater level of security than their home. This is a particularly insightful discovery and highlights just how essential mobile devices have become in our everyday lives. The rankings of each of the situations was found to be consistent across all of the age groups, with no real variation across different ages, suggesting that these are universal opinions (at least within our sample).

Following this, participants were asked which biometric technology that they would be comfortable using for each of the particular situations from the previous question. In order to determine which technology was perceived to be the most secure, amongst the respondents, analysis was carried out to understand the biometric that each participant perceived to be the most appropriate for the five situations, shown previously. For example, if a participant had ranked '*airport*' as the greatest need for security and then selected facial recognition as the appropriate technology for this situation, then this would be inferred to be the technology that the participant considered to be the most secure.

The results of this analysis can be seen in Figures 6 and 7, which shows distribution of biometrics selected as suitable for the most secure situation (Figure 6) and those that were selected as not suitable for any of the situations (Figure 7).

Firstly, focusing on those that were ranked as the most secure biometrics, it becomes apparent that there is a trend towards those methods that the participants were most familiar with. For example, the method most clearly perceived as the most secure was fingerprint recognition. This is something that is increasingly common in our daily lives, for example, the large majority

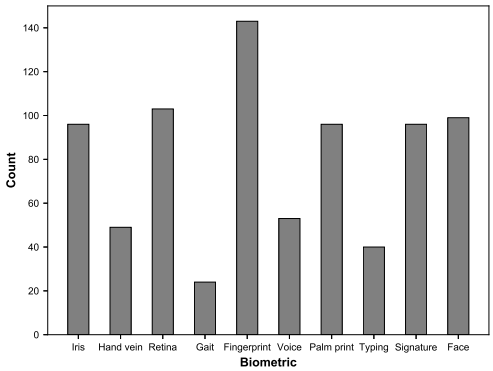

Figure 6: The biometrics that were ranked as the most secure

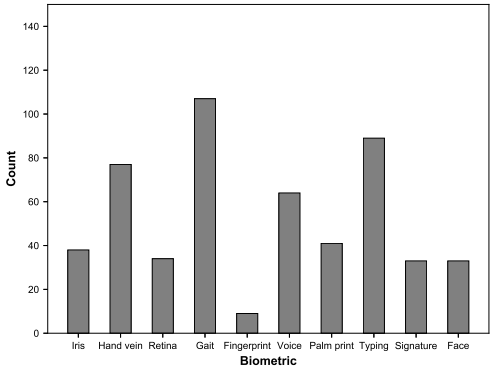

Figure 7: The biometrics that were ranked as the least secure

of personal devices (e.g., smartphones and tablets) now make use of fingerprint recognition and this is now becoming a common feature of international travel. Fingerprint recognition was also the biometric that the majority of participants had heard of and knowingly used, as seen in Figure 3, Figure 4 and Table 1. However, these fingerprint sensors in our personal devices are easily compromised, as demonstrated by both Yang et al. [27] and Kanchikere and Sudha [28]. This highlights that a lack of awareness of other biometric methods is potentially damaging to the personal security of users.

Another observation that can be made is that the biometrics that were perceived to be the most secure (iris, retina, palm print and facial recognition) are those methods that are perhaps either most commonly used in our day-to-day lives or those that are most common in popular culture. For instance, it is very common to see palm print recognition or retinal scans on our screens. However, in the case of facial recognition, as with fingerprint recognition, previous iterations of this technology has been shown to be easily compromised. For example, facial recognition has previously been circumvented with the use of printed masks/photos [29] [30].

At the other end of the scale it was noteworthy that the two behavioural biometrics (gait and typing analysis) were considered to be amongst the least secure. In addition to the two behavioural biometrics, voice recognition was also considered to be amongst the least secure. As previously discussed, three of these methods were all methods that had not been 'knowingly used' by the majority of participants and are all methods that could be used remotely without the user's knowledge. It is entirely possible that the intangible nature of these methods contributes to their perceived lack of security. Finally, hand vein recognition is amongst those method perceived to be the least secure. This is perhaps unsurprising given that this was the biometric that the fewest number of respondents had actually heard of, with none of the participants having knowingly used this method.

The final section of the survey asked participants for their opinions on the security of biometrics when compared to other authentication methods. First participants were asked whether they thought biometrics were as secure as passwords with 83% of participants agreeing that this was the case. Participants were then asked whether they thought that biometrics could provide the same level of security as two-factor authentication, with approximately 75% of people agreeing with this statement. Finally, participants were asked whether they felt that biometrics could be easily compromised. Only 46% of participants believed that biometrics were not easily compromised, which was a surprising result. This is especially true when considering that the majority of respondents felt that biometrics were as secure of two-factor authentication or passwords.

One of the key discoveries of this survey was that the participants generally felt that the methods they were most familiar with (e.g. fingerprints) were the most secure. This is perhaps not surprising but does highlight that familiarity exposure to these technologies helps to generate support for the methods.

4.2. Thematic and Word2Vec Analyses

To complement the research in Section 4.1, we were also keen to examine people's understanding of the meaning of the term *biometrics* itself. In our survey, we had asked participants whether they knew what was meant by the term *biometrics*, with 74% of respondents claiming to know what the term meant. Following this question these participants were asked to provide a definition of what they thought was meant by the term; this was completed by 49% of participants overall. This section analyses the responses that were given using two analysis techniques. First, we apply thematic analysis, which enables key themes to be discovered in the data [23]. After this, we explore the utility of an automated analysis approach called Word2Vec [24]; this technique models words in a vector space to allow for insight into textual content. Finally, we compare the findings of each technique to published definitions of biometrics today (e.g., [31] [32] [33]), which have been analysed using thematic analysis.

Definitions provide a simple yet effective method to elicit an individuals understanding of a particular term. We received a variety of definitions from participants for biometrics, which after analysis, resulted in several key themes. The most prominent of theme was that of *identity*, with biometrics being viewed as a means to identify, or verify the identity of, an individual. As one participant stated, it is the "use of biological/physical

parameters to identify a person". This therefore highlights a primary application of biometrics today according to individuals.

The second most significant theme also features in the quote above, i.e., the central part played by *biological and physical characteristics* within the identification process. According to participants, identification is thus not based on what a person knows (e.g., with passwords), it is driven by who they are, or in the words of another participant "[biometrics covers] essentially data about who you are/your body". These types of characteristics and data, along with the examples provided by participants (e.g., fingerprints, retina scans), match many of the types presented later in the study. Several participants even used examples of biological data as their definition for biometrics. While the general themes of identity and human characteristics were consistent, one participant made a point that biometrics allowed for identification "through non-traditional means". This view, although isolated, hints to a perception of biometrics as not yet fully streamlined as the new standard in identification.

Biometrics were also viewed as *a measure of* (or, statistics based on) key characteristics within the biological and physical space. To quote one participant, "a measure of the body using certain characteristics such as DNA or gait of walk that are individual to each person". This also highlights a *uniqueness* component, which enables biometric features to be somewhat distinguishing; a primary factor motivating their use in identification. Other noteworthy themes which emerged—albeit featuring significantly less—included *biometrics as behaviour*, and *biometrics for the purposes of security*. In the former theme, participants expressed that the behavioural characteristics of individuals were also central to the understanding of biometrics and how they are used. In most cases where behaviour was mentioned, it was alongside physical characteristics, for instance, biometrics are "related to people's physical and behavioural characteristics such as fingerprints, retinal scan, movement". In this case, physical characteristics were fingerprints and retinal scan, while the behavioural characteristic was movement (e.g., gait).

For the latter theme, biometrics were closely associated with security. As stated by a participant, biometrics were a means to "identify an individual by his unique biological identity / characteristic for security purposes". This pulls together several of the previous themes including identity, biological characteristics and uniqueness. One participant extended the role of security in biometrics further by describing it as "security measures based upon the human body". With another, it was perceived as "the use of body characteristics as an extra 'factor' in authentication". These views provide additional insight into how some individuals regard biometrics and topics including security, authentication and authorisation (mentioned in another quote).

After analysing the content manually using thematic analysis, we were interested in exploring the extent to which automated methods may be able to extract similar, or more nuanced, themes. This could also verify the findings of the manual analysis, or indeed, increase the confidence that may be placed in automated techniques in future analyses.

For this task, we decided to analyse the definitions using Word2Vec [24], an approach that models words in a vector space. This method takes a large corpus of text as its input, in this case the model was created using the Google News data (containing around 100 billion words), which includes 300-dimensional vectors for 3 million words and phrases [34]. This input is used to develop word vectors; first stop words (e.g., and, the, but) are removed, a vocabulary is constructed from the training text and then vector representations are identified for each word. Words can then be positioned within the vector space. Those words that are in close proximity to one another within the vector space share a common semantic meaning or context. A key point to note about this approach is that it derives context and relationships within the text based on the training corpus. As the corpus is created from publicly available text, it is free from biased on opinions and experiences of the researchers.

For our use of the method, we first took the biometrics definition from each individual and tokenised it into individual words, with stop words removed. Each of the remaining words was then projected onto vector space. The resulting word vectors were clustered, using k-means clustering [35] to determine the key concepts involved in the public perceptions of biometrics. It was determined, using the gap statistic approach [36], that the optimal grouping of the data would be into four clusters. This implied that the provided definitions could be distilled down into four key concepts. As k-means clustering is a stochastic approach it was repeated 1000 times. This resulted in 4 centroids in vector space per iteration. Each centroid can then be mapped back to find the work closest to its vector representation. These representations are shown in Figure 8 for the 1000 iterations.

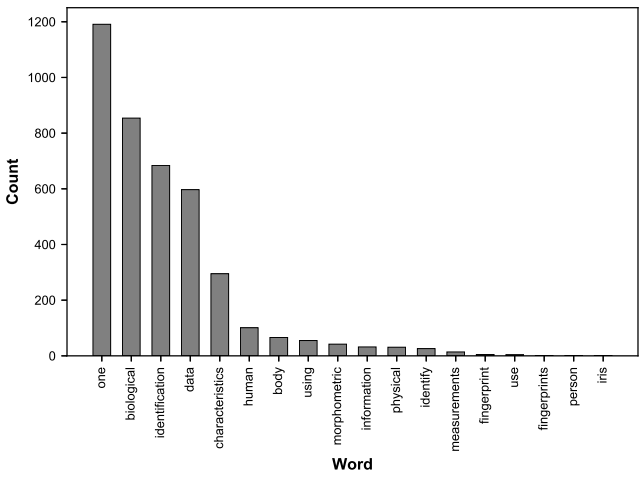

Figure 8: A graph of the most common word centroids based on the clustering approach

As can be seen in Figure 8, the five most common themes/terms that can be drawn out from the provided definitions are: one, biological, identification, data and characteristics. The most popular term, *one*, can be understood to refer to

oneself, or the individual. This could be taken to refer to the subject from whom biometrics may be attained. In the second and third terms, we can see the *biological* nature of biometrics being emphasised as well as their utility for *identification* purposes. The importance of *data* and identity *characteristics* to biometrics is highlighted in the two remaining prominent terms. These complement the previous themes and together allow for a clearer depiction of how participants understand and perceive modern-day biometrics.

To compare these findings with those from the manual analysis, there are several similarities. Most notably, the themes of identification, and biological and physical characteristics are prominent in both analyses. The starkest difference from a superficial perspective is the prominence of the term 'one' in the Word2Vec analysis. As discussed above however, this can be interpreted as relating to an individual i.e., the subject of the biometric. This would therefore align with the findings of the thematic analysis.

A less compensable difference in the two approaches is the importance of measures (as a theme) in the manual analysis, but not as high a ranking of measurements/metrics in the Word2Vec's clusters. This could be due to a significant number of unknown or new words. Word vector analysis, such as Word2Vec is known to struggle with words that are out of vocabulary (OOV), if the model has not encountered the word before then it will not know how build a vector representation. With this specific dataset there is a range of specialist vocabulary in use and as such may be OOV. Similarly, the accuracy of the vectorization process can be impacted if the Word2Vec model contains no shared representations at sub-word level. Word2Vec represents every word as an independent vector, despite many words being morphologically similar.

The next step in our research was to reflect on the perceptions and understanding of biometrics by our participants, as compared to the meaning of the term supplied by official and well-publicised sources. This would enable us to gauge how accurate—or rather, aligned—participants' perceptions were in light of standard definitions. For this task we used 16 sources: the four most well-known English dictionaries [32][37][38][39]; four standard-setting organisations [33][40][41][42]; three governmental organisations and a biometrics institute [43][44][45][46]; and four prominent texts [31][47][48][49]. These specific sources were selected because of either their nature, pedigree, or general need to be accessible to individuals of the public. After collecting definitions from each source, we applied thematic analysis (identical to our manual analysis of participants' data) to extract the core themes in the dataset.

From the completed analysis, we found a notable overlap in the themes arising from participant definitions and the official definitions. The most common themes in the official definitions were individuals or people who the biometrics would relate to, behavioural, biological and physical characteristics, and biometrics' use for identification and recognition. Merriam-Webster captures these themes aptly in defining biometrics as "the measurement and analysis of unique physical or behavioral characteristics (such as fingerprint or voice patterns) especially as a means of verifying personal identity" [38]. This description itself also highlights many of the themes that were identified in our participants' responses. As such, we might conclude that participants generally perceive and understand biometrics accurately, or at least that their understanding aligns with popular conceptualisations. There was one area where there was some disparity, however, i.e., behavioural characteristics and their importance to biometrics. In the official definitions, we found that behaviour featured in a majority of sources. If we consider participant responses however, behaviour was rarely discussed and there was a clear emphasis on biological or physical characteristics. This point further supports our survey results (as well as prior work by Furnell and Evangelatos [14]) regarding the lack of widespread awareness of behavioural biometrics.

This analysis of the individual definitions of the term *biometrics* has drawn out a number of interesting concepts and themes. It is particularly interesting to note that it would appear, as a result of this analysis, that the general public have a good grasp of most of the key areas surrounding the concept of biometrics. This with the exception of the behavioural component which we have discussed above.

5. Conclusions and Future Work

The survey has highlighted that there is seemingly a good level of awareness and acceptance of certain biometric methods. The participants had heard of the vast majority of methods, with only hand vein recognition being under represented, although this is perhaps understandable given that this technology is not particularly widespread. While there are significant differences between the underlying technologies of hand vein and palm print recognition the method of collection will look similar to the average person.

It is notable that while there is a general acceptance of these technologies, it is very much dependent on the context. However, the research has highlighted that users are seemingly the most comfortable with those methods that are more commonplace and familiar (e.g., fingerprint or facial recognition). When considering methods that are slightly more intangible (e.g., typing or gait analysis) they are typically less well regarded or understood. This suggests that perhaps there is scope to further develop the public's acceptance of these methods.

From our manual analysis of the definitions of biometrics provided by participants, we identified several key themes core to people's understanding of the topic. The complementary automated approach, Word2Vec analysis, was also shown to be able to highlight central themes in the definition data. This is commendable and could increase confidence in the approach's further use. Another key finding of our study is that, largely speaking, the definitions provided by participants closely aligned to the core elements of the more official definitions of biometrics. The exception to this being the behavioural component, which was also a result from our survey data.

Future work in this area would look to further develop a deeper understanding of the experiences and perceptions of biometrics technologies amongst the general public. This could be achieved by larger more representative samples, and cover

wider ranges of biometrics and related technologies which we were unable to address in this article; for instance, it could be interesting to investigate EEG and cardiac biometrics. Additionally, one of the key areas for future work is to establish methods for improving acceptance and understanding of the less tangible, behavioural biometric methods.

References

[1] B. Bajarin, Apple's Penchant for Consumer Security, 2016. Online: accessed 19 February 2019.

[2] Deloitte, Surge in UK adoption of fingerprint recognition points way to mainstream biometric authentication at the expense of the password, https://www2.deloitte.com/uk/en/pages/press-releases/articles/surge-in-uk-adoption-of-fingerprint-recognition.html, 2017. Online: Accessed 19 February 2019.

[3] BBC, Banks turning to voice recognition, 2016. Online: accessed 19 February 2019.

[4] The Telegraph, HSBC's voice recognition security breached by customer's brother, 2017. Online: accessed 19 February 2019.

[5] B. Grawemeyer, H. Johnson, Using and managing multiple passwords: A week to a view, Interacting with Computers 23 (2011) 256–267.

[6] Cyber Aware, Cyber Aware, 2017. Online: accessed 19 February 2019.

[7] R. Wash, E. Rader, R. Berman, et al., Understanding password choices: How frequently entered passwords are re-used across websites, in: Symposium on Usable Privacy and Security (SOUPS).

[8] E. F. Gehringer, Choosing passwords: security and human factors, in: Technology and Society, 2002.(ISTAS'02). 2002 International Symposium on, IEEE, pp. 369–373.

[9] SplashData, Worst Passwords of 2015, https://www.teamsid.com/worst-passwords-2015/, 2015. Online: Accessed 19 February 2019.

[10] RSA, RSA SecurID Hardware Tokens, 2017. Online: Accessed 19 February 2019.

[11] A. Jain, A. Ross, S. Pankanti, Biometrics: a tool for information security, IEEE transactions on information forensics and security 1 (2006) 125–143.

[12] J. Ashbourn, Biometrics in the new world: The cloud, Mobile technology and Pervasive identity, Springer Science & Business Media, 2014.

[13] J. D. Thompson, G. L. Herman, T. Scheponik, L. Oliva, A. Sherman, E. Golaszewski, D. Phatak, K. Patsourakos, Student misconceptions about cybersecurity concepts: Analysis of think-aloud interviews, Journal of Cybersecurity Education, Research and Practice 2018 (2018) 5.

[14] S. Furnell, K. Evangelatos, Public awareness and perceptions of biometrics, Computer Fraud & Security 2007 (2007) 8–13.

[15] K. J. Chan, S. J. Elliott, Privacy perceptions in biometrics operations, in: Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE), p. 143.

[16] M. Sabharwal, The assessment of concerns, opinions and perceptions of bankers to find the significant metrics for deployment of biometrics in e-banking, Biometrics and Bioinformatics 8 (2016) 27–38.

[17] C. Bhagavatula, B. Ur, K. Iacovino, Biometric authentication on iPhone and android: Usability, perceptions, and influences on adoption, Workshop on Usable Security (USEC) (2015).

[18] K. Krol, S. Parkin, M. A. Sasse, "I don't like putting my face on the Internet!": An acceptance study of face biometrics as a CAPTCHA replacement, in: Identity, Security and Behavior Analysis (ISBA), 2016 IEEE International Conference on, IEEE, pp. 1–7.

[19] M. Blum, L. A. Von Ahn, J. Langford, et al., The CAPTCHA project (Completely Automatic Public Turing test to tell Computers and Humans Apart), School of Computer Science, Carnegie-Mellon University, http://www. captcha. net (2000).

[20] O. Ogbanufe, D. J. Kim, Comparing fingerprint-based biometrics authentication versus traditional authentication methods for e-payment, Decision Support Systems 106 (2018) 1–14.

[21] D. Carpenter, A. McLeod, C. Hicks, M. Maasberg, Privacy and biometrics: An empirical examination of employee concerns, Information Systems Frontiers 20 (2018) 91–110.

[22] S. E. Cross, E. E. Hardin, B. Gercek-Swing, The what, how, why, and where of self-construal, Personality and Social Psychology Review 15 (2011) 142–179.

[23] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology 3 (2006) 77–101.

[24] T. Mikolov, K. Chen, G. Corrado, et al., Efficient estimation of word representations in vector space, in: International Conference of Learning and Representations (ICLR).

[25] L. A. Goodman, Snowball sampling, The annals of mathematical statistics (1961) 148–170.

[26] M. Williams, L. Axon, J. R. C. Nurse, S. Creese, Future scenarios and challenges for security and privacy, in: IEEE 2nd International Forum on Research and Technologies for Society and Industry (RTSI), IEEE, pp. 1–6.

[27] W. Yang, J. Hu, C. Fernandes, et al., Vulnerability analysis of iPhone 6, in: Privacy, Security and Trust (PST), 2016 14th Annual Conference on, IEEE, pp. 457–463.

[28] J. Kanchikere, R. Sudha, Hacking Mobile Phones Using 2D Printed Fingerprint, International Journal of Innovations & Advancement in Computer Science 7 (2018) 488–491.

[29] J. Galbally, R. Satta, Three-dimensional and two-and-a-half-dimensional face recognition spoofing using three-dimensional printed models, IET Biometrics 5 (2016) 83–91.

[30] SySS, Biometricks: Bypassing an Enterprise-Grade Biometric Face Authentication System, 2017. Online: accessed 19 February 2019.

[31] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, A. W. Senior, Guide to biometrics, Springer Science & Business Media, 2013.

[32] OED, Biometrics, n.d. Online: Accessed 19 February 2019.

[33] ISO, ISO/IEC 2382-37:2017 Information technology – Vocabulary – Part 37: Biometrics, 2017. Online: Accessed 19 February 2019.

[34] T. Mikolov, I. Sutskever, K. Chen, Distributed representations of words and phrases and their compositionality, in: Advances in neural information processing systems, pp. 3111–3119.

[35] J. A. Hartigan, W. A. Wong, Algorithm AS 136: A k-means clustering algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (1979) 100–108.

[36] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (2001) 411–423.

[37] CUP, Biometrics, n.d. Online: Accessed 19 February 2019.

[38] Merriam Webster, Biometrics, n.d. Online: Accessed 19 February 2019.

[39] Collins, Biometrics, n.d. Online: Accessed 19 February 2019.

[40] NIST, Biometrics Testing, 2009. Online: Accessed 19 February 2019.

[41] ITU, Biometrics and Standards, 2017. Online: Accessed 19 February 2019.

[42] BSI, Biometrics British Standards, 2018. Online: Accessed 19 February 2019.

[43] UK Home Office, Biometrics Strategy: Better public services, Maintaining public trust, 2018. Online: Accessed 19 February 2019.

[44] DHS, Biometrics, 2017. Online: Accessed 19 February 2019.

[45] EU, Art. 4 GDPR Definitions, 2018. Online: Accessed 19 February 2019.

[46] Biometrics Institute, Biometrics, 2017. Online: Accessed 19 February 2019.

[47] R. J. Anderson, Security engineering: a guide to building dependable distributed systems, John Wiley & Sons, 2008.

[48] W. S. Coats, A. Bagdasarian, T. Helou, T. Lam, The practitioner's guide to biometrics, American Bar Association.

[49] A. Jain, P. Flynn, A. A. Ross, Handbook of biometrics, Springer Science & Business Media, 2007.