

The value of data: Estimating the value individuals put on access to their computer files

Anna Cartwright ^{*} Edward Cartwright [†] Lian Xue^{‡§¶}

June 2021

Keywords: Data, Cyber-security, cyber-crime, willingness to pay, willingness to accept

Abstract

Cyber attacks are a growing threat that undoubtedly cause significant loss to individuals. We know, however, surprisingly little about the true economic cost of such attacks. In this paper we estimate the value individuals put on access to their computer files by eliciting willingness to pay (WTP) and willingness to accept (WTA) from a representative sample of the UK population. We report the results of three surveys, which vary in the methods used to elicit WTP and WTA, and give broadly consistent results. We find that the distribution of valuations is highly skewed to the right, with a median WTP of around £50-1000 and a median WTA of £700-1000. We find a large and systematic difference between WTP and WTA which may reflect the ‘psychological’ cost of cyber-crime.

^{*}School of Economics, Finance and Accounting, Coventry University, UK.

[†]Department of Economics and Marketing, De Montfort University, UK. Corresponding author, email edward.cartwright@dmu.ac.uk.

[‡]School of Economics and Management, Wuhan University, China.

[§]This project has received funding from the Engineering and Physical Sciences Research Council (EPSRC) for project EP/P011772/1 on the Economic, Psychological and Societal Impact of Ransomware (EMPHASIS). EC also wants to thank the European Union’s Horizon 2020 research and innovation programme, under grant agreement No.700326 (RAMSES project), which also supported this work.

[¶]We would like to thank four anonymous reviewers for WEIS2021 for their comments and advice on an earlier version of the paper.

1 Introduction

There is no doubt that cyber crime is now a significant threat to individuals. For instance the 2020 crime figures in England and Wales show that the annual number of recorded victims from computer misuse cases (876,000) was higher than that of burglary (582,000), robbery (127,000) car theft (80,000) and criminal damage (323,000).¹ Some may argue that losing access to digital files is less costly (economically and psychologically) than, say, the theft of a car or domestic burglary. But, the simple truth is, we don't have the data to make that judgment. Indeed, we have surprisingly little understanding of the costs of cyber-crime. Instead, the media and literature are full of numbers, typically very large, that are of questionable accuracy (Anderson et al., 2013; Armin et al., 2015). This makes it difficult to quantify cyber risk relative to other risks. It also provides difficulties for policy makers trying to discern how much resource to devote to cyber-security.

As a concrete example, consider ransomware. This is a financially motivated crime in which files are encrypted and a ransom is demanded for the key to decrypt the files (Kharraz et al., 2015; Maigida et al., 2019). A widely quoted estimate of the cost of ransomware in 2017 was \$5 billion.² In 2018 and 2019 the estimates had risen to \$8 billion and \$11.5 billion.³ The underlying methodology behind deriving these estimates is, however, entirely unclear. A number of academic studies have traced bitcoin payments in order to provide a precise lower bound on the size of ransomware payments made to criminals (e.g. Liao et al., 2016; Bursztein et al., 2017; Huang et al., 2018; Paquet-Clouston et al., 2019). The resultant estimates are in the range of \$13 to \$26 million (in total) between 2013 and 2017. The costs of ransomware clearly extend well beyond the payment of ransoms. Even so, the huge gap between these precisely estimated numbers and those widely quoted in the field is a concern.

To provide more reliable estimates on the cost of cyber-crime we need to break the problem up into manageable chunks (Anderson et al., 2013; Home Office, 2018). In this paper we focus on the cost of cyber-crime to individuals. We have relatively reliable estimates on the amount individuals spend on mitigating attack through virus protection or similar, and the amount lost through well recorded cyber crimes such as fraud (Home Office, 2018). We do not, however, have much understanding of the cost to individuals of losing access to their digital files. Our objective in this paper is to put a monetary estimate on the amount individuals value access. We also look at how this value varies with individual characteristics, such as age and employment status, as well as with attitudes to and knowledge of cyber-security. Let us emphasize that our focus will be on an individuals access to their own files, rather than data breach or data theft whereby others gain access to those files. We can, thus, give insight on malware, of which ransomware is one type, that destroys or encrypts data.

To meet our objective we obtain data (from a representative sample of the UK population) on willingness to pay (WTP) and willingness to accept (WTA) for recovery of

¹These are ONS Crime Statistics for 2020 available at <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/crimeinenglandandwalesappendixtables>. A computer misuse case is defined as unlawful access to a computer and device and so includes hacking and viruses. This does not include online fraud.

²See <https://cybersecurityventures.com/ransomware-damage-report-2017-5-billion/>

³See <https://www.riskiq.com/press-release/evil-internet-minute-1138888-lost-cybercrime-reveals-riskiq/> and <https://www.acronis.com/en-gb/articles/costs-of-ransomware-attacks/>

computer files. Elicitation of WTP and WTA has a long history of use in economic contingent valuation of intangible or non-monetary assets. For instance, it is widely used in valuing environmental, health and cultural assets (Olsen and Smith, 2001; Throsby, 2003; Alberini and Kahn, 2006). Contingent valuation can also prove useful in measuring the value of privacy or the non-monetary costs of cyber-crime (Acquisti et al., 2013; Hernandez-Castro et al., 2020). In our context, we assume that an individual has lost access to their files because of, for example, a ransomware attack. WTP is the maximum amount of money the individual is willing to pay to restore access to their files. WTA is the minimum amount of money the individual would be willing to receive (from some third party) in-lieu of restored access to their files. For instance, if the individual’s WTA is £200 then they would rather have £210 than their files, and would rather have their files than £190. WTP and WTA provide, therefore, measures of how much people value access to their files.

We report the results of three separate surveys in which we vary the framing of the key WTP and WTA elicitation questions. Results are broadly consistent across the three surveys giving confidence in our approach. We find that the distribution of valuations is highly positively skewed. Moreover, there is a large discrepancy between WTP and WTA. These two factors make it difficult to put an aggregate measure on the cost of data loss. For instance, the median WTP is around £50-100 while the mean WTA is over £10,000. The cyber-security industry would presumably want to focus on the latter number but we will suggest that median and modal values of WTA, in the range of £700-£1000, are probably more informative. There is, though, considerable heterogeneity across the sample. For instance, anything up to a half of individuals have a low valuation (near £0) and over a tenth have a high valuation (of over £1,000).

An important question is what drives heterogeneity in valuations. A number of factors could enter the mix including the type and amount of files, the sensitivity of those files and the time since the last back up. For instance, a low valuation could mean the individual simply has no valuable files, or that they had valuable files but their security behaviour, such as regular back-ups, diminishes the loss from missing files on a particular device. On the flip side, a high valuation could mean some deficiency in security, because a back-up does not exist, and/or ‘psychological cost’ from the loss of sensitive information. To explore these issues we obtained data on a range of variables, including back-up frequency, knowledge of cyber risk, attitudes to data breach and knowledge of cyber-security best practice.

We find that these variables interact in an interesting way with WTP and WTA. To illustrate one manifestation of the effect (the more subtle effects will be discussed later), individuals who have concerns about data breach or are knowledgeable of cyber-security best practice tend to have a relatively high WTA and low WTP. This seems intuitive if we think of WTA in terms of an individual wanting ‘compensation for psychological costs’ of losing files. Such effects help explain why we observe a large gap between WTP and WTA. They also suggest that care is needed in interpreting WTP and WTA because they may be measuring subtly different things. Indeed we will use such results to argue that WTA is a better measure of the cost of cyber-crime.

As far as we are aware the only other study that has elicited WTP and WTA for computer files is Hernandez-Castro et al. (2020). They report the results of a small-scale pilot survey (149 participants) that compares different ways of wording the WTP and WTA questions. We build on that approach here by using a larger, more

representative, sample and also collecting a raft of additional socio-economic and cyber related data. Our raw data on WTP and WTA is similar to that of Hernandez-Castro et al. (2020) although we obtain higher values for WTA.

A related literature uses WTP and WTA to elicit how much individuals value personal information (e.g Huberman et al., 2005; Grossklags and Acquisti, 2007; Carrascal et al., 2013; Dogruel et al., 2017; Wagner et al., 2018). One interesting finding from this literature, in connection with our work, is a large discrepancy between WTP and WTA (Acquisti et al., 2013). In drawing, however, a connection between our work and that on privacy we highlight a difference between own access to files, which is what we are studying in this paper, and third party access to files, which is relevant for privacy. An individual may, for instance, be willing to pay to restore access to sentimental photos but have no concerns about others viewing those files. Alternatively, an individual may not be willing to pay to restore sensitive work files, because there is a ready back up, but be willing to pay to avoid those files being leaked.

We finish this introduction by noting that any study based on contingent valuation can be criticised on the grounds that stated preferences can differ from revealed preferences (Winegar and Sunstein, 2019; Hausman, 2012). Indeed, a large literature has explored the privacy paradox, the notion that stated and revealed preferences differ in disclosing private information (Kokolakis, 2017). We would argue, however, that this is not a reason to discount contingent valuation studies, particularly in a setting where there is a dearth of information on which to draw. It does, though, mean that conclusions need to take account of potential reasons why stated and revealed preferences may differ. For instance, the privacy paradox may result from immediate gratification bias in which an individual discloses private information for less than would be optimal because of impatience (Acquisti, 2004). Hence, the individual sells information for less than stated WTP. In our setting, such bias could mean an individual pays more to restore access to files than would be optimal. Hence, they pay more than stated WTP.

We proceed as follows. In Sections 2, 3 and 4 we detail the methods and results of studies 1, 2 and 3 respectively. In Section 4 we conclude.

2 Study 1

In our first study we elicited WTP and WTA from a large representative sample of the UK population. We also obtained basic data on demographic characteristics and computer use. This survey provides, therefore, an overall picture of the distribution of WTP and WTA across the UK population.

2.1 Methods

We ran a survey on a representative sample of the UK population. The survey was run using YouGov. A total of 2002 participants were surveyed. The participants were 53% female, 12% aged 18-24, 12% aged 25-34, 14% aged 35-44, 17% aged 45-54 and 45% aged 55 or over. Participants were geographically dispersed across the UK with 62% in social classification ABC1. Also, 52% were in work, 46% married and 23% with children in the household.

We used the following questions to measure willingness to pay (WTP) and willingness to accept (WTA) for loss of files:

WTP: Please imagine that you accidentally lost access to ALL the files (e.g. photographs, documents) on your main personal digital device (e.g. computer, laptop, tablet, smartphone). The only way to recover your files is to pay a private company who are experts in file recovery. Approximately, how much is the MAXIMUM you would pay in order to restore access to ALL your files on your device? (Please type your answer in the box below to the nearest pound "£". If you are unsure, please give your best estimate)

WTA: Please imagine that your main personal device (e.g. computer, laptop, tablet, smartphone) has been infected by a virus, which means you have lost access to ALL the files (e.g. photographs, documents) on your device. The criminals responsible for the virus have been caught and you are eligible for monetary compensation. Approximately, how much is the MINIMUM you would consider fair compensation for the loss of ALL your files on your device?

The WTP and WTA questions were asked at distinct points during the survey to obtain independent answers. Moreover, the ordering of questions was varied so that half of the participants were exposed to the WTP question first and half to the WTA question first. We find no discernible difference due to question ordering (Mann Whitney test, $p < 0.01$). The wording of the questions (e.g. pay a private company or criminals responsible) was informed by (Hernandez-Castro et al., 2020) who find no significant effect on WTP or WTA due to different framings. We, therefore, chose frames that we believed would be easy for participants to relate to. In Surveys 2 and 3, reported below, we consider alternative framings.

We asked participants what type of files they store on the device with answers including photographs (79% of participants), music (47%), movies (15%), work and education documents (29%), personal documents e.g. notes (48%), and sensitive documents e.g. financial information (24%). We asked questions about security measures including frequency of back-up (23% backing-up every week). We asked whether they know of ransomware (63%) and online identity theft (86%) and whether they had experienced ransomware (4.5%). We also asked how concerned they are with a data breach with answers ranging from not concerned at all (7%), to not very concerned (30%), to fairly concerned (44%) and very concerned (14%). We also have data on use (within the previous month) of social media platforms including Facebook (70%), Twitter (29%), Instagram (22%) and WhatsApp (44%).

In answering our key WTP and WTA questions participants had the option to say 'Do not know' or 'Unwilling to answer'. While this can reduce noise in the data, because participants are not forced to 'come up with a number', it significantly reduced the number of observations we will analyse. Specifically, 1006 participants answered the WTP question, 968 the WTA question and 807 answered both questions. Table 1 summarises the sample. In terms of sample selection bias the set of people answering the WTP and WTA questions are significantly different from the overall sample ($p < 0.001$, joint orthogonality test). In particular, participants were significantly more likely to answer the questions if they were male and young. You can see in Table 1, however, that the sample of people answering the WTP and WTA questions are still broadly representative of the overall sample. We shall, therefore, proceed on the basis that there is no large self-selection bias.

Table 1: Representativeness of sample for WTP, WTA and both WTP and WTA questions.

	Overall	WTP	WTA	P&A
Male	0.46	0.54	0.54	0.55
Age	46.9	45.7	45.5	45.2
Married	0.46	0.44	0.45	0.44
Have children	0.24	0.23	0.23	0.23
Working full-time	0.37	0.40	0.39	0.40
Social grade ABC1	0.61	0.66	0.68	0.69
<i>Sample Size</i>	2002	1006	968	807

2.2 Results

Figures 1 and 2 plot the distribution of valuations for WTP and WTA, respectively, across all observations. Both the WTP and WTA are highly influenced by a positive skew. The median WTP and WTA is £50 and £700, respectively, and the modal WTP and WTA is £0 and £1000. A discrepancy between WTP and WTA is to be expected (Kahneman et al., 1991; Sayman and Öncüler, 2005; Alberini and Kahn, 2006) and in environmental studies the difference can be anything from a factor of two to ten (Horowitz and McConnell, 2002). In our case we observe a difference towards the higher end of this range. That poses a conundrum when we come to put a figure on the cost of cyber-crime. Should we go with WTP or WTA? And should we go with the median, mode or mean? The figure we get for the loss of files is clearly going to be wildly different depending on what we choose. This means that it is important for us to dig deeper, as we shall now do, and question in more detail the factors that are influencing the WTP and WTA.

We have seen that there is a large heterogeneity in valuations. To explore in more detail the factors that influence an individual’s WTP and WTA, we report the results of a tobit regression with the log of WTP or WTA as the dependent variable (or more formally $\log(WTP + 1)$ and $\log(WTA + 1)$). Table 2 provides the results. We can see that WTP is significantly higher for females and those using LinkedIn. It is decreasing in age. WTP is also significantly higher if the individual is storing work and/or sensitive documents on the device. This is consistent with the data on privacy valuation (Acquisti et al., 2013; Carrascal et al., 2013). In terms of data breach, we see an inverse U-shaped relationship in which those very worried about data breach and those not worried about data breach have a lower WTP.

There are some notable differences between WTA and WTP. With WTA we still see a significant positive effect of storing sensitive documents and using LinkedIn, and a negative effect from not being worried about data breach. Interestingly, though, we see females have a lower WTA, and WTA is increasing with age. In interpretation, this means that men have a larger gap between WTP and WTA than women. Put differently, men have a relatively low WTP and high WTA when compared to women. Similarly, older people have a larger gap between WTP and WTA. This most likely reflects different responses to the framing of the questions. For instance, it may be that men and older individuals expect a larger compensation from criminals and, thus,

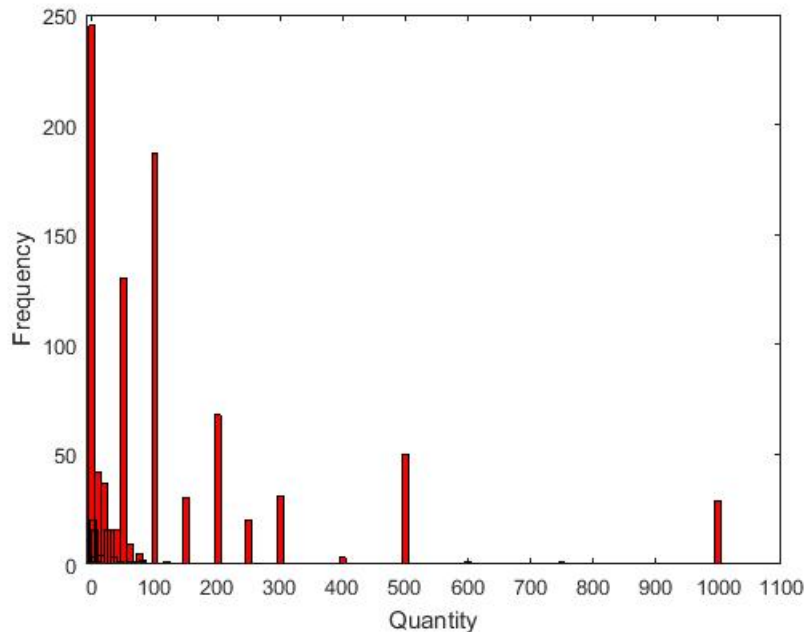


Figure 1: Histogram of WTP for Survey 1. Notable peaks at £0, £50, £100, £200, £500 and £1000. We only plot WTP less than £1100.

report a higher WTA.

2.3 Discussion

Survey 1 shows proof of concept that we can elicit WTP and WTA for loss of access to files. One notable feature of the data is extreme heterogeneity in valuations. This is to be expected given the inevitable differences in how much people use and rely on digital technology. Even so, it suggests that we cannot easily talk of an ‘average victim’ of cyber-crime. Broadly speaking, around a half of participants reported a low WTP, of below £50. Around a third report a mid WTP, of between £100-300. That leaves around a sixth of participants who report valuations in excess of £500. We have seen that women, young people and those that store sensitive and work documents are more likely to have a high valuation.

Another notable feature of our results is a large gap between WTP and WTA. While some gap is to be expected the gap we observe is extreme. Indeed, a majority of individuals expressed a WTA of £500 or above. This suggests that WTA and the notion of ‘compensation’ is not just measuring the cost of losing access to files but also the cost from being attacked and/or a willingness to exploit the option to over-claim compensation. As ‘psychological’ compensation for criminality may be relevant if we wish to measure the cost of cyber-crime. We need, however, to investigate in more detail why the WTA values are so much higher than WTP. In Surveys 2 and 3 we amend the framing of the WTA question to remove the notion of compensation and of criminality.

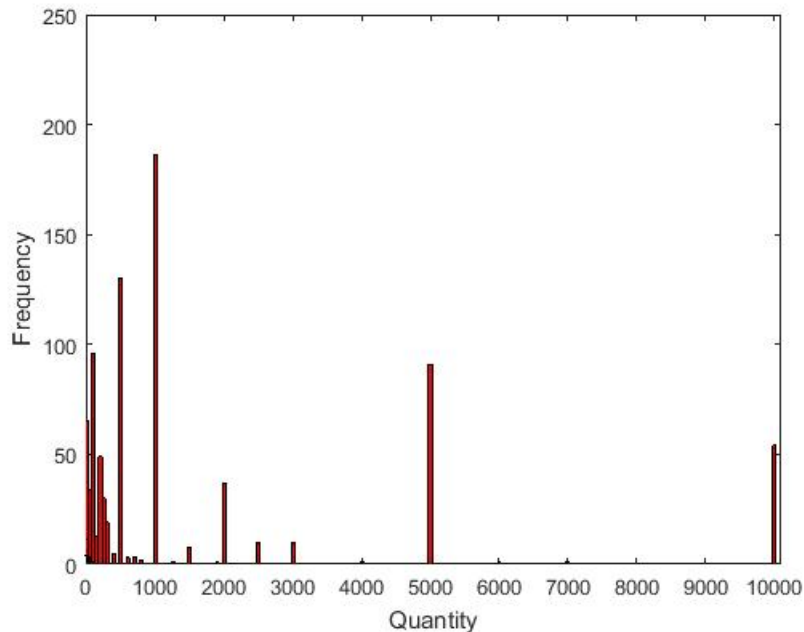


Figure 2: Histogram of WTA for Survey 1. Notable peaks at £100, £500, £1000, £5000 and £10000. We only plot WTA less than £10100.

A large literature has looked at gender differences in preferences, attitudes to risk and over-confidence (Eckel and Grossman, 2002; Bengtsson et al., 2005; Croson and Gneezy, 2009). A separate literature has documented gender differences in cyber-security behaviour (Anwar et al., 2017; Gratian et al., 2018). It is, therefore, not surprising to see gender differences in WTP and WTA. The effect we see, however, is primarily one of women being more consistent across WTP and WTA than men. Gender differences in the WTP WTA disparity have been observed before (Wieland et al., 2014).

3 Study 2

In Survey 2 we revise the WTA elicitation question to avoid any notion of criminality or compensation. We also obtain more detailed data on the individual’s cyber-security approach and attitudes. This allows us to explore if there is a positive correlation between the amount an individual values their files and the security measures in place to protect those files. To reduce the heterogeneity of the sample we focused exclusively on individuals who were currently in employment.

3.1 Methods

A total of 404 participants, resident in the UK and in employment, took part in a survey. Participants were recruited using Prolific. The participants were 65% female, 37% aged

Table 2: Tobit regression of WTP/WTA. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

VARIABLES	(1) WTP	(2) WTA
Constant	1.185 (0.209)***	3.26 (0.225)***
Female	0.177 (0.084)**	-0.175 (0.092)*
Age	-0.117 (0.033)***	0.073 (0.036)**
Social_grade C2DE	0.116 (0.085)	0.082 (0.093)
Working	0.155 (0.080)*	0.030 (0.087)
Married	0.121 (0.091)	0.129 (0.099)
Children	1.01 (0.098)	-0.271 (0.107)**
Store photos	0.105 (0.115)	0.027 (0.127)
Store music	0.115 (0.107)	0.039 (0.095)
Store work	0.317 (0.089)***	-0.057 (0.118)
Store sensitive	0.346 (0.091)***	0.212 (0.100)**
Back up every week	-0.011 (0.092)	0.097 (0.099)
Know ransomware	-0.165 (0.092)*	0.001 (0.101)
Know online theft	0.068 (0.139)	0.390 (0.147)***
Experienced ransomware	-0.226 (0.187)	-0.467 (0.196)**
Use Facebook	0.085 (0.092)	-0.46 (0.100)
Use Twitter	-0.023 (0.090)	0.053 (0.098)
Use LinkedIn	0.247 (0.112)**	0.233 (0.118)**
Very worried data breach	-0.179 (0.089)**	0.108 (0.128)
Not worried data breach	-1.045 (0.164)***	-0.303 (0.098)***
Not at all worried data breach	-1.05 (0.164)***	-0.698 (0.183)***
<i>Observations</i>	1006	968
<i>Pseudo R²</i>	0.052	0.04

18-29, 35% aged 30-39 and 27% aged 40 or over. There were 33% of participants earning less than £30,000/year, 44% earning between £30,000-60,000/year and 22% earning £60,000/year or above. Also, 62% were in full-time employment, 34% married and 43% with children.

As previewed, the wording of the key WTP and WTA questions were slightly different to the of Study 1 in order to test the robustness of our findings to variants in framing. Note that WTA is now framed in terms of being offered money for an experiment rather than compensation for criminality.

WTP: We now want you to imagine that you accidentally lost access to all the files (e.g. photographs, documents, data) on your main personal device. You have no back-up. The only way to recover your files is to pay a private IT company who are experts in file recovery. Approximately how much (to nearest £) is the maximum you would pay in order to restore access to all your files?

WTA: We next want you to imagine that (as part of an experiment) you are offered money to delete all of the files from your personal device. You have no back-up and

would have no chance of recovering your files once deleted. What is the minimum amount of money (to the nearest £) you would accept to delete your files?

We asked participants how often they use digital technology with answers on a 0-6 scale from never (0) to every hour. We separately asked about frequency of using email (average response of 4.99), social media (4.69), pictures/video (3.50), watching movies/listening to podcasts (3.41), online banking (3.51), purchasing goods (2.89), word processing/spreadsheets (3.22) and games (2.65). Participants were then asked to think ‘about the personal digital device (mobile phone, laptop, etc.) you use most frequently for storing important files?’ They were asked what this device was, with answers of smartphone (38.4%), tablet (4.5%), laptop (42.1%) and desktop (12.9%). As in Study 1 participants were asked whether they store on the device photos (88%), music (51%), movies (23%), work (58%), personal files (77%) and sensitive files (46%).

As well as asking participants how often they back up their device we asked them whether they use anti-virus (69%), firewall (47%), password manager (47%), 2 factor authentication (61%), automatic back-up (45%), cloud storage (67%) and a location tracking app (37%). To measure beliefs and self-confidence we asked participants how confident they are that they can spot a phishing attack. Responses were measured on a 0-3 scale from not at all confident to very confident (average response 1.9). They were also asked on a 0-5 scale how familiar they were with firewalls, anti-virus, virtual private network (VPN), and 2 factor authentication. We combined these four measures to create an aggregate score of familiarity with security measures (average score 3.52).

Finally, we asked participants two questions to probe their knowledge of cyber-security best practice. They were asked which of five passwords was considered most secure: 43ghy78!, rangeTotalcavern, Cart00nwaveBuilding, 4936582, and ManchesterUntd. The ‘correct answer’, based on National Cyber Security Centre guidance, was rangeTotalcavern or Cart00nwaveBuilding. Only 14% chose this response.⁴ Participants were also asked what the government currently advises people to do when browsing on a public network (e.g. train station or hotel): do not use public networks at all, use a VPN, only use sites that are password protected, or only use secure https websites. The ‘correct answer’ being use a VPN (with 26% correct). We say that those who got both questions correct are familiar with best practice.

3.2 Results

Figures 3 and 4 plot the distribution of valuations for WTP and WTA respectively across all observations. The distribution in both cases are broadly similar to that from Study 1. In terms of WTP we no longer see a modal peak at 0. We still, though, see a cluster of observations around £0 and peaks at £100, £500 and £1000. Similarly, we again see large peaks in WTA at £1,000, £5,000 and £10,000. As in Study 1 the distribution of WTP and WTA is highly influenced by positive skew. The median WTP and WTA is £100 and £5000 respectively. The modal WTP is £100 and £10,000 respectively.

The discrepancy between WTP and WTA is even higher in Survey 2 than survey 1. In particular, the median value of WTA minus WTP is £450 in Survey 1 but rises to

⁴The NCSC guidance is to use three random words and so we have checked the robustness of our results to also allowing rangeTotalcavern as a correct answer. Only 1% of participants chose this option. Including this as a correct answer does not change our results.

£4900 in survey 2. The discrepancy between WTA and WTP we observed in Survey 1 does not, therefore, appear to result from basing the WTA elicitation question around compensation or criminality.

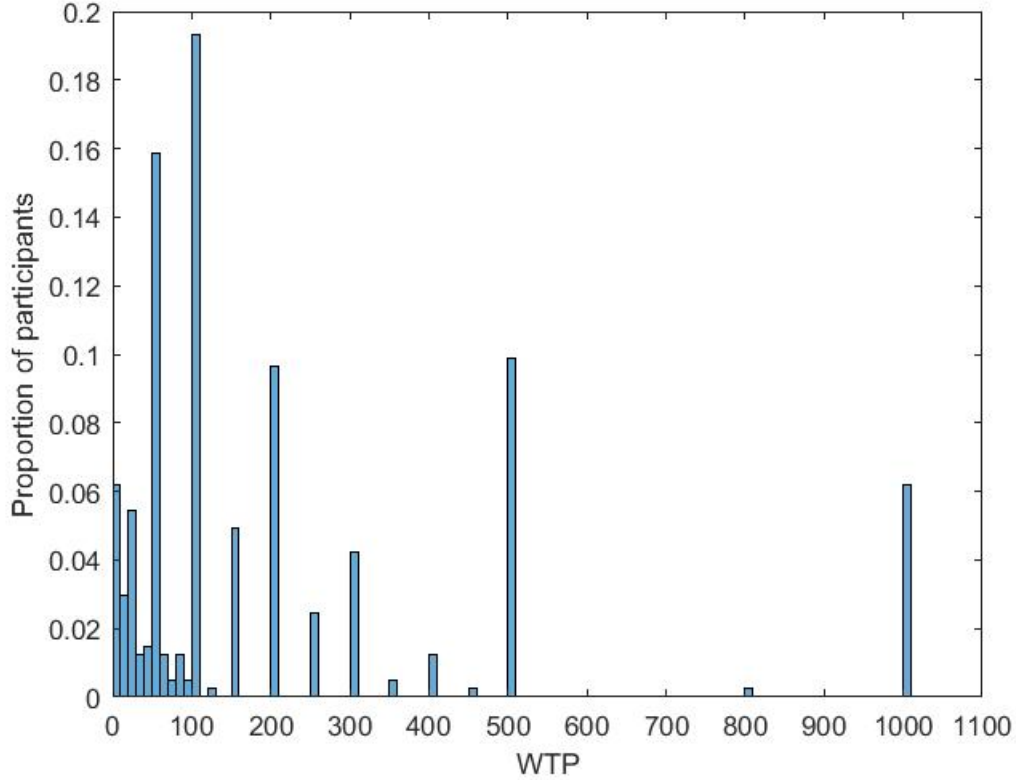


Figure 3: Histogram of WTP in Survey 2. Notable peaks at £50, £100, £200, £500 and £10000. We only plot WTP less than £1100.

We performed a tobit regression with $\log(WTP + 1)$ and $\log(WTA + 1)$ as the dependent variable. Table 3 provides the results. We find that WTP is significantly higher for those storing photos. Unlike Survey 1 we find no effect from work or sensitive documents. This is primarily because we now control for use of digital technology and use of word processing, in particular, is highly correlated with WTP and storage of work or sensitive documents. Clearly, increased time spent word processing is likely to increase the number of work and sensitive files. We find that those who know cyber-security best practice had a significantly lower WTP. Those using a desktop had a significantly higher WTP (compared to the benchmark of using of laptop).

In terms of WTA we find a strong positive effect of knowledge of best practice. In interpretation, this means that those with knowledge have a relatively low WTP and high WTA. We find that WTA is strongly increasing with use of email and decreasing in familiarity with cyber-security. We do observe a significant relationship between WTA and storing of sensitive documents. Unlike Survey 1, we find a strong negative effect of age on WTA.

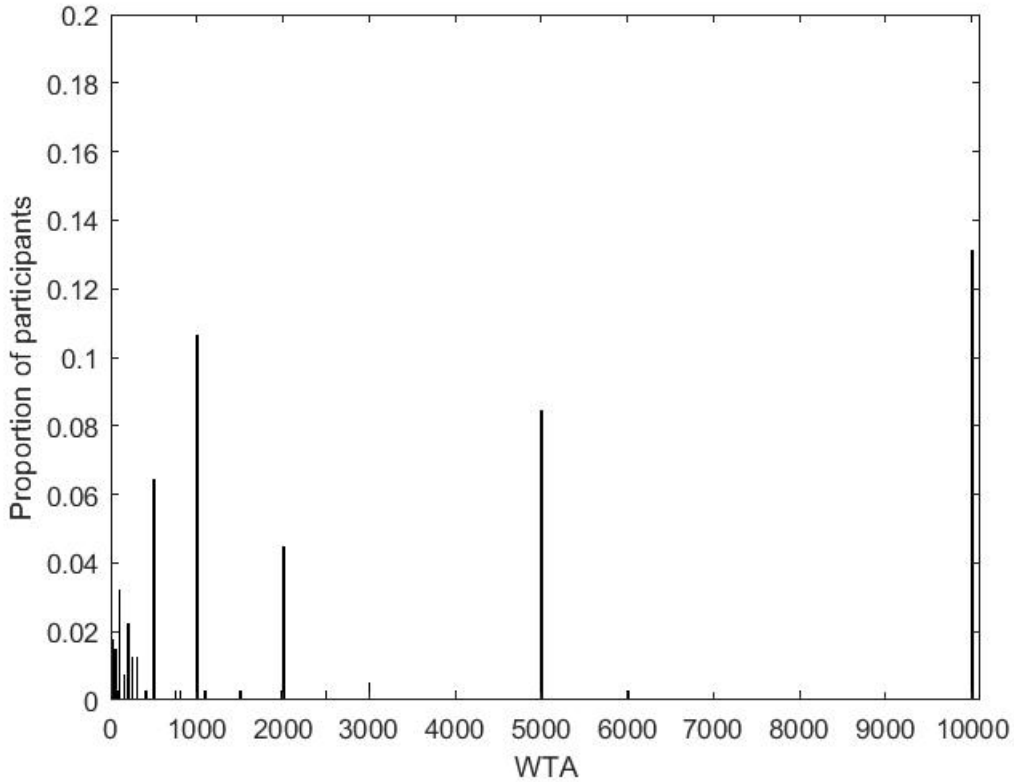


Figure 4: Histogram of WTA in Survey 2. Notable peaks at £500, £1000, £2000, £5000 and £10000. We only plot WTA less than £10100.

3.3 Discussion

Survey 2 broadly confirms the results of Survey 1 in terms of the distribution of WTP and WTA. Indeed, the distributions are remarkably similar. The elicited values of WTP and WTA are slightly higher in Survey 2 but that is not unexpected given that in Survey 2 we focused on individuals in employment.

One key finding from Survey 2 is that the discrepancy between WTP and WTA remains, despite framing the WTA elicitation question in a way that does not mention criminals or compensation. This lends support to the notion that the high WTA captures a ‘psychological cost’ from loss of files that is independent of criminality. This could reflect a cost, including in time and effort, of ‘destroying’ files (and resistance to a ‘pointless task’).

4 Study 3

In Studies 1 and 2 we asked participants to state a willingness to pay and willingness to accept for loss of access to files. This is arguably a difficult task for individuals to do because it is a relatively abstract question. An alternative way of measuring WTP

Table 3: Tobit regression of WTP/WTA for Study 2. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

VARIABLES	(1) WTP	(2) WTA
Constant	0.816 (0.326)**	2.171 (1.031)**
Female	0.051 (0.093)	0.046 (0.294)
Age	-0.045 (0.039)	-0.293 (0.122)**
Income	0.029 (0.052)	0.145 (0.163)
Graduate	0.010 (0.084)	0.271 (0.263)
Married	0.093 (0.097)	-0.209 (0.317)
Children	0.017 (0.098)	0.489 (0.308)
Store photos	0.293 (0.123)**	0.289 (0.386)
Store music	0.125 (0.085)	0.364 (0.269)
Store work	0.062 (0.094)	-0.245 (0.295)
Store personal	0.156 (0.101)	0.018 (0.316)
Store sensitive	-0.010 (0.085)	0.524 (0.266)**
Back up frequency	0.061 (0.033)*	0.140 (0.104)
Security measures	-0.019 (0.026)	0.134 (0.081)
Phishing confidence	0.056 (0.047)	0.158 (0.147)
Familiarity with cyber	-0.065 (0.039)*	-0.254 (0.122)**
Knows best practice	-0.0380 (0.188)**	1.55 (0.590)***
Use of email	0.065 (0.053)	0.442 (0.169)***
Use of videos	0.064 (0.041)	0.013 (0.129)
Use of movies	0.025 (0.037)	-0.057 (0.117)
Use of games	-0.043 (0.026)*	0.002 (0.082)
Use of social media	-0.001 (0.034)	-0.187 (0.107)*
Use of banking	0.019 (0.043)	0.054 (0.138)
Use of retail	0.006 (0.053)	-0.173 (0.168)
Use of word processing	0.062 (0.029)**	0.042 (0.090)
Smartphone	-0.127 (0.101)	-0.466 (0.317)
Tablet	-0.194 (0.196)	-0.329 (0.614)
Desktop	0.302 (0.125) **	0.186 (0.392)
<i>Observations</i>	404	404
<i>Pseudo R²</i>	0.074	0.028

and WTA is to ask individuals if they be willing to pay or accept a specified amount of money for the loss of files (Sayman and Öncüler, 2005; Alberini and Kahn, 2006). While this method does not provide the full distribution of WTP and WTA across the population it can give more reliable estimates around the specified amount. As a robustness check we, thus, carried out a third survey using this elicitation method. We compared specified amounts of £100, £300 and £500.

4.1 Methods

A total of 219 participants, resident in the UK and in employment, took part in a survey. Participants were recruited using Prolific. The participants were 74% female, 41% aged 18-29, 32% aged 30-39 and 27% aged 40 or over. Participants were 38% earning less than £30,000/year, 41% earning between £30,000-60,000/year and 21% earning £60,000/year or above. Also, 60% were graduates, 32% married and 41% with children.

The design of Survey 3 closely follows that of Survey 2 except the key WTP and WTA questions are revised around a specified amount. A total of 116 participants were asked the following questions:

WTP: Imagine that you accidentally lost access to all the files (e.g. photographs, documents, data) on your main personal device. You have no back-up. The only way to recover your files is to pay a private IT company who are experts in file recovery. The company charges £300. Would you pay this?

WTA: Now imagine that (as part of an experiment) you are offered money to delete all of the files on your main device. You have no back-up and would have no chance of recovering your files. You are being offered £300 to take part in the experiment. Would you take part?

A total of 52 subjects were asked the same questions but the amount revised to £100. And 51 subjects were asked about £500. Participants answered on a 0-3 scale from definitely no, probably no, probably yes and definitely yes.

Participants use of digital services and security measures was very similar to Survey 2. We measured frequency of use of email (average response of 4.82), social media (4.76), pictures/video (3.48), watching movies/listening to podcasts (3.58), on-line banking (3.43), purchasing goods (2.92), word processing/spreadsheets (3.02) and games (2.71). We elicited the device they use most often for storing files, including smartphone (43.4%), tablet (3.2%), laptop (39.2%) and desktop (10.5%). We elicited whether they store on the device photos (86%), music (48%), movies (20%), work (58%), personal files (81%) and sensitive files (48%). We also elicited the use of anti-virus (57%), firewall (32%), password manager (42%), 2 factor authentication (58%), automatic back-up (42%), cloud storage (66%) and location tracking app (39%).

Participants expressed their confidence in spotting a phishing attack on a 0-4 scale from not at all confident to extremely confident (average response 2.6). Familiarity with security measures was scored as in Survey 2 (with average score of 2.92). Knowledge of cyber-security best practice was also measured as in Survey 2 with 13% 'correct' on the password question and 31% 'correct' on the public wi-fi question.

4.2 Results

In Table 4 we summarise responses to the WTP elicitation question and compare the results with those of Survey 1 and Survey 2. The results for Surveys 1 and 2 are based on the proportion of participants who indicated a WTP at or above £100, £300 and £500 respectively. As one would expect we see that there is less willingness to pay £500 than £300 than £100. The comparison with surveys 1 and 2 is particularly interesting. If we focus purely on those participants who said definitely yes (i.e. 3) then we obtain a lower proportion willing to pay than in surveys 1 and 2. If, by contrast, we focus on those who said probably or definitely yes (2 or 3) then we find a higher proportion willing to pay than in Surveys 1 and 2. It would seem reasonable, therefore, to conclude that the ‘true’ proportion is somewhere between those answering definitely yes and those answering probably or definitely yes. We note that the higher estimates of WTP in Surveys 2 and 3 (compared to Survey 1) are expected given that surveys 2 and 3 have a higher proportion of women and focus on people in employment (see results of Study 1).

Amount	Survey 3					Survey 1	Survey 2
	0	1	2	3	2 + 3		
£100	9.6	15.4	48.8	26.9	75	43.2	63.4
£300	12.7	35.3	40.5	12.1	52.6	12.8	26.7
£500	13.7	49.1	33.3	3.9	37.3	9.6	20.5

Table 4: Proportion of participants (%) willing to pay to recover files, where responses range from 0 definitely no, 1 probably no, 2 probably yes to 3 definitely yes.

In Table 5 we summarise responses to the WTA elicitation question. The results for surveys 1 and 2 are based on the proportion of participants who indicated a WTA below £100, £300 and £500 respectively. You can see that we observe a relatively low willingness to accept in Survey 3. Indeed, only 15.7% indicate a willingness to accept £500 for loss of their files. If we base willingness to accept on the combination of probably yes or definitely yes then the results of Study 2 are very similar to those of Survey 2. This, again, reinforces the point that the high WTA values of Study 1 are not due to the framing being around compensation for criminality. It also reinforces the point that elicited WTA is considerably higher than WTP.

In Table 6 we report the results of an ordered probit regression with WTP and WTA as the dependent variable. As one would expect we see a significantly lower willingness to pay £500 and higher willingness to pay £100 compared to the baseline of £300. We see that WTP is significantly lower for those who are married and significantly higher for those who store photos. There is a positive relationship between WTP and frequency of back up and between WTP and use of videos. There is a negative relationship between use of games and WTP, and between familiarity with cyber and WTP. The only factors that are significant in terms of WTA are familiarity with cyber and use of games. Both of these factors show a reverse sign compared to WTP meaning that these factors are associated with a larger gap between WTP and WTA.

Amount	Survey 3					Survey 1	Survey 2
	0	1	2	3	2 + 3		
£100	75	19.2	3.8	1.9	5.8	14.2	10.9
£300	58.6	23.3	11.2	6.9	18.1	33.6	18.3
£500	43.1	41.2	11.8	3.9	15.7	36.1	19.8

Table 5: Proportion of participants (%) willing to accept to recover files, where responses range from 0 definitely no, 1 probably no, 2 probably yes to 3 definitely yes.

Table 6: Ordered probit regression of WTP/WTA for Study 3.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

VARIABLES	(1)	(2)
	WTP	WTA
£100	0.629 (0.201)***	-0.593 (0.232)**
£500	-0.476 (0.197)**	0.196 (0.207)
Female	-0.263 (0.202)	0.139 (0.219)
Age	-0.108 (0.083)	-0.001 (0.091)
Income	0.159 (0.103)	-0.030 (0.114)
Graduate	0.180 (0.166)	-0.094 (0.181)
Married	-0.406 (0.198)**	-0.175 (0.220)
Children	0.175 (0.197)	-0.121 (0.218)
Store photos	0.865 (0.250)***	0.067 (0.262)
Store music	-0.148 (0.166)	-0.248 (0.181)
Store work	-0.043 (0.181)	-0.007 (0.197)
Store personal	0.052 (0.223)	-0.143 (0.242)
Store sensitive	-0.142 (0.173)	0.049 (0.190)
Back up frequency	0.252 (0.064)***	-0.072 (0.069)
Security measures	-0.033 (0.054)	-0.056 (0.058)
Phishing confidence	-0.070 (0.094)	0.018 (0.103)
Familiarity with cyber	-0.148 (0.074)**	0.242 (0.082)***
Knows best practice	-0.598 (0.362)*	-0.072 (0.410)
Use of videos	0.182 (0.085)**	-0.138 (0.094)
Use of games	-0.106 (0.050)**	0.125 (0.056)**
Use of banking	0.048 (0.109)	-0.015 (0.124)
Use of retail	-0.063 (0.121)	0.074 (0.130)
Use of word processing	0.081 (0.055)	-0.019 (0.062)
Smartphone	-0.394 (0.199)**	0.276 (0.217)
Tablet	0.499 (0.468)	-0.418 (0.564)
Desktop	0.553 (0.301)*	-0.224 (0.339)
<i>Observations</i>	219	219
<i>Pseudo R²</i>	0.151	0.091

4.3 Discussion

In terms of the distribution of WTA and WTP, the results of Study 3 are consistent with those of Studies 1 and 2. This gives confidence in our findings. In terms of the factors that influence WTP, we see some consistent results across Studies 2 and 3. In particular, there is a positive relationship between back up frequency and WTP in both surveys. Similarly, there is a negative relationship between knowledge of best practice and WTP, and use of games and WTP in both surveys. WTP is lower for those using smartphones and higher for those using desktops, compared to laptop. There is a strong positive relationship between WTP and storage of photos. Finally, the frequency of use of games is negatively related with WTP in both surveys. In terms of the factors that influence WTA, we see less consistency across Studies 2 and 3. For instance, familiarity with cyber and WTA is positive on Survey 2 and negative on Survey 3.

5 Concluding discussion

Our basic objective in this paper was to put some science behind the valuation of individual losses from cyber-crime. We know that cyber-crime leads to loss of files and data but how much value do individuals put on that loss? Answering this question is essential in evaluating the overall costs of cyber-crime. We use self-reported WTP and WTA for loss of files as a measure of the value individuals put on those files. One can criticize such self-reporting measures as ‘cheap talk’. But, to put things bluntly, it is better than nothing, which is what we have now. We believe self-reported WTP and WTA is a step in the right direction in measuring loss from cyber-crime and we hope that future work will expand on and develop our approach. We have reported the results of 3 surveys which give consistent results, as summarized in Table 7.

One important finding in our work is a large heterogeneity in valuations of files with a large positive skew. One question this raises is whether we should be using mean or median to approximate aggregate societal loss. We would strongly suggest the median is more appropriate because the mean is heavily influenced by some large values (Florencio and Herley, 2013). This is not to say that there are no individuals who would indeed be heavily effected by loss of files. But to take account of this it would be better to capture the distribution of loss rather than focus on a large mean.

Broadly speaking we can distinguish three or four categories of individual. If we focus on willingness to pay, then around a half of individuals have a low WTP of £0-50, a third of individuals have a medium WTP of £100-500, and a sixth with a high WTP. If we focus on willingness to accept, then around a sixth of individuals have a low WTA of £0-50, around a quarter have a low-medium WTA of £100-1000, a quarter have a medium-high WTA of £1000-5000, and a third with a higher WTA. The exact distribution, though, clearly depends on a range of factors including socio-economic factors and use of digital technology.

One key finding is a large discrepancy between WTP and WTA. We would argue that this discrepancy follows from differences in the interpretation of WTP and WTA in our setting. In particular, WTP is a measure of how much an individual would pay to return access to files while WTA has an element of compensation for loss of files. In one scenario the person gets the files back and the other they do not. This difference, seems

	Study 1	Study 2	Study 3
WTP £100	43.2%	63.4%	26.9-75%
WTP £300	12.8%	26.7%	12.1-52.6%
WTP £500	9.6%	20.5%	3.9-37.3%
Median WTP	£50	£100	
Mode WTP	£0	£100	
WTA £100	14.2%	10.9%	1.9-5.8%
WTA £300	33.6%	18.3%	6.9-18.1%
WTA £500	36.1%	19.8%	3.9-15.7%
Median WTA	£700	£5000	
Mode WTA	£1000	£10000	

Table 7: Summary of key points of distribution of WTP and WTA from the 3 studies

to drive a wedge between WTP and WTA. A large discrepancy between WTP and WTA has also been observed for data privacy (Acquisti et al., 2013) and is surely related. A critical question to explore, therefore, is whether WTP or WTA (as measured) is most appropriate for measuring the loss incurred from cyber-crime.

There are general arguments to suggest that the true valuation will be closer to WTA than to WTP (Bateman et al., 2005). Here we would argue that there are additional factors pointing towards the merits of WTA. Specifically, we find that WTA correlates with variables, such as concerns about data-breach, knowledge of cyber-security best practice and familiarity with cyber-security, that suggest WTA may capture better the psychological costs of loss of files. To motivate the point consider someone who is highly concerned about data breach and very knowledgeable and familiar with cyber-security. Our analysis suggests that this person will not have a high WTP, presumably because their ‘good cyber behaviour’ allows them to recover from the loss of files. But this person does have a high WTA which may capture the additional psychological cost from having been attacked (even if the direct material loss is small).

If pushed to come up with a number we would, therefore, suggest the median WTA is a good measure of the average cost from loss of files. This gives a number of £700 per individual in Study 1. The median WTA in Study 2 is much higher (at £5000) but this was obtained with a sample likely to have higher incomes and higher use of digital technology. Moreover, 37% of individuals had a WTA of £1000 or less. An average WTA in the region of £700-1000 seems, therefore, justifiable. Our study is only the first step on the way of more accurately characterizing the losses from cyber-crime. We need more surveys and more understanding of the factors that influence WTP and WTA. We also need to feed in issues around vulnerability of attack. For instance, those with the highest WTA may be less vulnerable to attack because of good security measures. To measure the overall loss to individuals from cyber-crime we, thus, need to match loss with risk exposure. Indeed, a critical finding from our analysis is that cyber awareness significantly decreases an individuals willingness to pay a ransom. Individuals will, therefore, be exposed to different risk irrespective of how much they value their files.

References

- A. Acquisti. Privacy in electronic commerce and the economics of immediate gratification. *Proceedings of the 5th ACM conference on Electronic commerce*, pages 21–29, 2004.
- A. Acquisti, L. K. John, and G. Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274, 2013.
- A. Alberini and J. Kahn. *Handbook on contingent valuation*. Edward Elgar Publishing, 2006.
- R. Anderson, C. Barton, R. Böhme, R. Clayton, M. JG Van Eeten, M. Levi, T. Moore, and S. Savage. Measuring the cost of cybercrime. In *The Economics of Information Security and Privacy*, pages 265–300. Springer, 2013.
- M. Anwar, W. He, I. Ash, X. Yuan, L. Li, and L. Xu. Gender difference and employees’ cybersecurity behaviors. *Computers in Human Behavior*, 69:437–443, 2017.
- J. Armin, B. Thompson, D. Ariu, G. Giacinto, F. Roli, and P. Kijewski. 2020 cyber-crime economic costs: No measure no solution. *2015 10th International Conference on Availability, Reliability and Security*, pages 701–710, 2015.
- I. Bateman, D. Kahneman, A. Munro, C. Starmer, and R. Sugden. Testing competing models of loss aversion: An adversarial collaboration. *Journal of Public Economics*, 89(8):1561–1580, 2005.
- C. Bengtsson, M. Persson, and P. Willenhag. Gender and overconfidence. *Economics letters*, 86(2):199–203, 2005.
- E. Bursztein, K. McRoberts, and L. Invernizzi. Tracking desktop ransomware payments. *Black Hat USA, Las Vegas, United States*, 2017.
- J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira. Your browsing behavior for a big mac: Economics of personal information online. *Proceedings of the 22nd international conference on World Wide Web*, 2013.
- R. Croson and U. Gneezy. Gender differences in preferences. *Journal of Economic literature*, 47(2):448–74, 2009.
- L. Dogruel, S. Joeckel, and J. Vitak. The valuation of privacy premium features for smartphone apps: The influence of defaults and expert recommendations. *Computers in Human Behavior*, 77:230–239, 2017.
- C. C Eckel and P. J Grossman. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4):281–295, 2002.
- D. Florencio and C. Herley. Sex, lies and cyber-crime surveys. *Economics of Information Security and Privacy III*, pages 35–53, 2013.
- M. Gratian, S. Bandi, M. Cukier, J. Dykstra, and A. Ginther. Correlating human traits and cyber security behavior intentions. *Computers & Security*, 73:345–358, 2018.

- J. Grossklags and A. Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. *Workshop on the Economics of Information Security*, 2007.
- J. Hausman. Contingent valuation: from dubious to hopeless. *Journal of Economic Perspectives*, 26(4):43–56, 2012.
- J. Hernandez-Castro, E. Cartwright, and A. Cartwright. An economic analysis of ransomware and its economic consequences. *Royal Society Open Science*, (190023), 2020.
- Home Office. Understanding the costs of cyber crime - a report of key findings from the costs of cybercrime working group. 2018.
- J. K. Horowitz and K. E. McConnell. A review of wta/wtp studies. *Journal of Environmental Economics and Management*, 44(3):426–447, 2002.
- D. Y. Huang, D. McCoy, M. M. Aliapoulos, V. G. Li, L. Invernizzi, E. Bursztein, K. McRoberts, J. Levin, K. Levchenko, and A. C. Snoeren. Tracking ransomware end-to-end. In *Tracking Ransomware End-to-end*. IEEE, 2018.
- B. A. Huberman, E. Adar, and L. R. Fine. Valuating privacy. *IEEE Security & Privacy*, 3(5):22–25, 2005.
- D. Kahneman, J. L. Knetsch, and R. H. Thaler. The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206, 1991.
- A. Kharraz, W. Robertson, D. Balzarotti, L. Bilge, and E. Kirida. Cutting the gordian knot: A look under the hood of ransomware attacks, 2015.
- S. Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64:122–134, 2017.
- K. Liao, Z. Zhao, A. Doupé, and G-J. Ahn. Behind closed doors: measurement and analysis of cryptolocker ransoms in bitcoin. In *Electronic Crime Research (eCrime), 2016 APWG Symposium on*, pages 1–13. IEEE, 2016.
- A. M. Maigida, M. Olalere, J. K. Alhassan, H. Chiroma, and E. G. Dada. Systematic literature review and metadata analysis of ransomware attacks and detection mechanisms. *Journal of Reliable Intelligent Environments*, 5(2):67–89, 2019.
- J. A. Olsen and R. D. Smith. Theory versus practice: a review of ‘willingness-to-pay’ in health and health care. *Health Economics*, 10(1):39–52, 2001.
- M. Paquet-Clouston, B. Haslhofer, and B. Dupont. Ransomware payments in the bitcoin ecosystem. *Journal of Cybersecurity*, 5, 2019.
- S. Sayman and A. Öncüler. Effects of study design characteristics on the wta–wtp disparity: A meta analytical framework. *Journal of Economic Psychology*, 26(2): 289–312, 2005.

- D. Throsby. Determining the value of cultural goods: How much (or how little) does contingent valuation tell us? *Journal of Cultural Economics*, 27(3):275–285, 2003.
- A. Wagner, N. Wessels, P. Buxmann, and H. Krasnova. Putting a price tag on personal information—a literature review. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
- A. Wieland, J. Sundali, M. Kemmelmeier, and R. Sarin. Gender differences in the endowment effect: Women pay less, but won’t accept less. *Judgment & Decision Making*, 9(6), 2014.
- A. G. Winegar and C. R. Sunstein. How much is data privacy worth? a preliminary investigation. *Journal of Consumer Policy*, 42(3):425–440, 2019.