COMS30038 — SECURITY BEHAVIOURS

SOCIAL ENGINEERING: A THEORY OF PERSUASION

Matthew Edwards[*]

---

1 WHY COMPUTER SCIENTISTS NEED TO KNOW ABOUT SOCIAL ENGINEERING

Computer security is hard because, unfortunately, people want to use their computers. This inconvenient fact has been the bane of many otherwise excellent security technologies. The *confidentiality* of some information stored on an electronic device can be assured through the judicious application of encryption and/or fire, as thoroughly destroyed devices tell no secrets. Also specifying that we would like to protect the *integrity* of such information forces the security engineer to stop gleefully torching hardware, but can be managed with various message authentication and digital signature schemes. It is only when the user dares to suggest that they should also be able to access this information – that they need *availability* as well as both of the above properties – that the problem starts to get serious.

"How am I going to keep something confidential," the long-suffering security engineer might ask, "if you want me to show it to a person? How do I make sure it's not meddled with, if you want me to let somebody edit it?"

At first blush, she might sound a little ridiculous. After all, military and intelligence organisations were protecting secrets for millennia before computers came along. A security policy can, no doubt, apply to the behaviour of humans as well as to computers and communication systems. If a policy is correctly designed, and people stick to it, you can make assurances about the security of a system. Properly educating people about the importance of sticking to the policy — perhaps by threatening them with severe penalties if they don't — then handles the human side of the problem, just like good cryptography and software engineering handles the machine side. Proper models of where information should and should not be going are of course critical to both parts of such a design, defining who (or what) is or isn't allowed to do this or that.

[*]matthew.john.edwards@bristol.ac.uk

However, the problem the security engineer is thinking of is this:

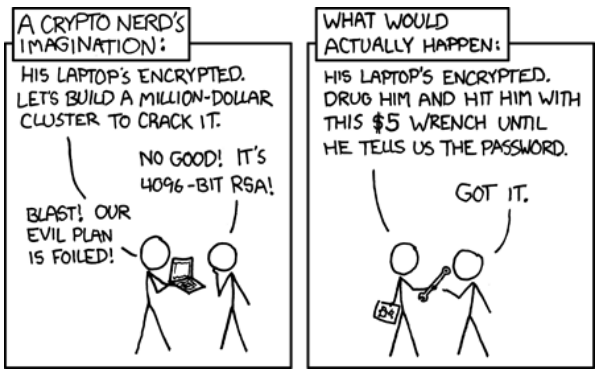

Figure 1: XKCD 538: Security [1]

There is no leverage that any organisation — private or public — can bring to bear on a person so that they will *never* deviate from a policy. There are always pressures which can be brought to bear on a person which will outweigh the noncompliance penalties you have threatened. Torture is just one extreme of a continuum of such pressures, most of which are situational and psychological, and any of which can cause people to do things regardless of whether or not their actions are permitted by a security policy. This is independent of additional concerns about whether the user properly understands and remembers what it is they are supposed to do according to the policy. In many cases, as will be discussed in this course, a person may not even be consciously aware they are violating a security policy.

These human vulnerabilities are important considerations for anyone trying to secure a system that involves humans, and are often the things that make security engineers break out in a cold sweat. However, in this part of COMS30038 we are interested first of all in how people *break* security. The art of exploiting these human vulnerabilities is referred to as *social engineering*, and it is one of the most important skills for any computer criminal.

1.1 *The Classics: Social Engineers in History*

The con can be a timeless art. Eugéne François Vidocq[1] in his *Memoirs*, reproduces an example of the sort of letter he recalls seeing prisoners craft in the Bicêtre in the late 18th century. The letter opens by addressing the recipient's presumed surprise at this letter from a person unknown to them, then proceeds to outline a tale of sorrow and injustice, related to current events, which concludes with the opportunity for the recipient to profit greatly by lending their assistance to the author [2, p. 58]. That summary, though recounting

[1] Vidocq was a larger-than-life character, both a criminal famed for his skill at fraud, disguises and jailbreaks, and "the father of modern criminology", the first private detective, and the founder of the French National Police Force. In the very first chapter of his autobiography he recounts (among other exploits) how he deserted several times *from both sides* of the War of the First Coalition. Of course, he was also a magnificent liar, so caveat lector.

letters handwritten in French back before the invention of the ballpoint pen, could just as well describe many modern emails being delivered via Russian botnets from Nigerian cyber-cafes. The diamonds of a late marchioness are replaced with the bank account of a wealthy businessman, and the scattered conflict of the French Revolution is updated to a plane crash, but the format remains identifiably intact. You'll get the chance to study and write letters like this during this week's lab session.

Another classical tactic is impersonation – the pedigree of impersonation and disguise as a means of deceit goes back to Homer, if not further. Frank Abagnale Jr.[2], who spent much of his criminal career forging cheques, was also a great impersonator, posing variously as an airline pilot, a doctor, a teacher, a prison agent, and a lawyer. He relates one impersonation episode where he noticed the location where airlines and car rental businesses dropped off their daily intake at an airport drop-box. Disguised as a security guard using a uniform that he had purchased at a local costume shop, he left a sign at the drop-box which read, "Out of Service. Place deposits with security guard on duty." Abagnale sat there while businesses brought him money. Later, he would express disbelief this idea had actually worked. After all, "How can a drop-box be out of service?" [3, p. 159]. We'll cover pervasive modern forms of digital impersonation later this week.

Moving into the digital age, one of the most famous computer criminals of the 20th century, Kevin Mitnick, was known to rely heavily on social engineering. Starting out as a phreaker[3], Mitnick quickly picked up the lingo of telephone engineers, learning how to talk like one of them so he didn't arouse suspicion when he rang up asking for details he needed to carry out his phreaking. These soft skills, and the ability to learn what he needed to say to sound natural in various different settings, soon became his single most transferable asset when he moved on to computer intrusion:

> "I could often get passwords and other pieces of sensitive information from companies by pretending to be someone else and *just asking for it*." — Kevin Mitnick [4, p. 21].

Like many hackers, Mitnick didn't work alone. One of his companions, Susan Headley (a.k.a. Susan Thunder) was another famed social engineer, and a real-life *femme fatale*, known to seduce high-ranking military officers in order to search their personal effects for passwords and access codes. When their other companion Lewis de Payne (a.k.a. Roscoe) spurned Headley for another woman, she set out on a campaign of revenge against him and Mitnick, which quickly became a kind of hacker war – each side struggling for the upper hand across bulletin boards, phone lines and paper trails (both literal and

[2] Best known as the real-life criminal inspiration behind the book, film and Broadway musical *Catch Me If You Can*, Abagnale spent five years in prison for his crimes before being released to help the FBI combat fraud – something he still does today.

[3] *Phreaking* is an early hacker culture that focused on reverse-engineering the public telephone system – and in particular the frequencies used to permit free long-distance calls.

electronic). Headley won this war, turning over a pile of carefully-accrued evidence that saw de Payne and Mitnick arrested, in return for her own immunity from prosecution [5, p. 15–61]. Despite having only an eighth-grade education, Headley later became a famous security consultant who would testify before Congress. You may not be surprised to learn that she was also a professional poker player [6].

There is a common thread through each of these individuals' stories. All of them, from Vidocq to Headley, were at one point criminals, and later 'turned cop' and started working with the authorities to catch those like themselves. Why were all these organisations — from the Parisian police to the FBI — so eager to adopt these ex-cons? Simple. They had the skills. They knew how to ply a con, and so they knew how to catch someone else in the act. A week's instruction cannot hope to compete with a lifetime of criminal experience, but the aim of this part of COMS30038 is to give you a glimmer of that ability – to show you how to carry out a scam, so you can recognise one when you're being targeted.

As a final example, let's discuss Kane Gamble. In 2015, a 15-year-old British teenager managed to hack, among others, the US Secretary of Homeland Security, the chief of the CIA and the deputy director of the FBI, obtaining a number of very sensitive documents about operations in Afghanistan and Iran. How did he achieve this? He rang customer services. In particular, in the case of Mr. Brennan, the chief of the CIA, he rang Verizon customer services, pretending first of all to be an employee, so he could understand how they operate. Then he rang back pretending to be Mr. Brennan, managing to convince the call handler to change his pin and security questions. From there he moved on to Brennan's AOL account, accruing details and logins until he was able to access all of his emails, contacts, cloud storage, and even his wife's iPad [7, 8]. This technique was incredibly powerful. At one point during his breakdown of the security of the FBI deputy director Mark Giuliano, the FBI cottoned on that someone had breached their online portal, and changed the password. In response, Gamble called their helpdesk and got them to change it again so he could get back in [8]. Very helpful!

This wasn't the result of any radical innovation. These techniques are basic social engineering, something someone like Vidocq would have found natural. In fact, the exact playbook Gamble was following was described in great detail by Mitnick in 2003 [4, p. 34–38]. Yet, more than a decade later, the United States' topmost security professionals were still so vulnerable that they were humiliated by a British teenager.

1.2 *The Scale of Social Engineering*

It can be hard to get reliable figures for the number of attacks which involve social engineering, but the consensus certainly seems to be 'a lot'. In 2003 the FBI estimated[4] that 80% of attacks involve social engineering. More recent

[4] In an online report which no longer exists and might be apocryphal.

estimates[5] range as high as 97 or 98%. If you ask hackers themselves, 84% of them say they use social engineering[6] – but the ratio between attackers and (successful) attacks would probably not be 1:1. One of the most completely described analyses comes from Verizon's annual Data Breach Investigations Report[7], which in its 2020 issue reports that of all 3,950 breaches during the 2019 observation period, 22% involved the use of social engineering. The majority of this social engineering activity was phishing for credentials, which was overall the most common single tactic in all breaches [9, p. 13].

Of course, this is just reported data breaches – these attacks target the general population of web users as well, for whom credential loss is often unknown or unreported. This is a growing issue. In the first quarter of 2020, 165,772 unique phishing sites were observed [10, p. 7] over a period of three months. In the last quarter of 2020, there were at least 190,000 unique phishing sites detected *every* month [11, p. 3]. Fraud targeted at the general public is growing in scale at a significant rate, and the methods are diversifying.

2 METHODS OF INFLUENCE

The core element of social engineering is persuasion – talking someone into doing something for you. The 'something' they do might be anything from giving you the name of someone you want to speak with to resetting an account password or opening a locked door. We are going to start out by looking at how persuasion works psychologically, from a high-level perspective. The aim here is to give you some tools to think about persuasion before you jump into the more practical end of carrying this out (the focus of the next lecture, and the labs).

The "useful model"[8] we can refer to here is known as the *elaboration likelihood model*, first formulated by the psychologists Richard Petty and John Cacioppo back in the 1980s [12]. The core idea of this model is there are two routes to persuasion. The first is the 'central' or 'cognitive' route to persuasion. This is where you persuade someone by arguing a case – by presenting a series of statements that, if accepted, should logically entail that the person follows the course of action you intend. A prototypical example might be a scientific paper, where the authors want to convince you to believe something about the nature of reality, and they present an argument based on the results of experiments. If you accept both the interpretation of their results and the structure of their argument, you should agree with their conclusion.

The second route is the 'peripheral' or 'non-cognitive' route. This is where

[5] Provided by cybersecurity companies, who often have no information about how they arrived at these figures, but are happy to sell you some software for your protection.

[6] Based on a sample of 70 DEFCON attendees, who might not be particularly representative of cybercriminals generally – we'll talk more about the difficulty of studying cybercrime in some sessions later in the course.

[7] `https://enterprise.verizon.com/resources/reports/dbir/`

[8] As the popular saying goes, all models – especially those of phenomena as complex as human behaviour – are wrong or incomplete, but many of them can be useful.

you persuade someone not with the content of your argument, but with peripheral cues – how you say it, who you are, what your relationship with this person is. This is a more subtle form of persuasion, one that might affect your judgement without you really being aware that it is happening, or what is causing you to change your mind[9].

Though these are described as two distinct routes, in reality they can be viewed more like complementary strategies, and they are often seen together. Consider a typical television advert for a product – let's say a car. The aim of the advertiser is, clearly, to persuade you to buy the car. The advert will make a series of claims about the benefits of this car — it can automatically park for you, it is fuel-efficient, it will protect you well in case of an accident, etc. — these are, at least in part, central persuasion. There may be a compelling case in there that you should buy the car, because it would be in your own self-interest. However, there are is almost certainly peripheral persuasion going on in the advert as well. If a person appears in the advert, they are likely to be unusually physically attractive. They perhaps, with the help of the car, attract the attention of other physically attractive people. Some of them might be famous. The advert is likely to be set in a remote area of outstanding natural beauty, and the car will drive down roads that contain no other vehicles. The advert might attempt to suggest that purchasing the car is something particularly noble or impressive people would do. These are all indirect suggestions, most of which would sound ridiculous if spelled-out: there is no mechanism for the car to affect how you look; it's not going to make you famous; buying the car will not teleport you to a beautiful stretch of isolated countryside and allow you to live there; buying the car does not itself make you noble or impressive. Yet for a lot of people these suggestions may be having an impact on their judgement of whether they should buy that car.

Of course, this effect is not limited to advertising and cybercrime. Persuasion is an ordinary component of human interaction, and we are constantly exerting these influences on each other. One reaction to learning about persuasion strategies is that you start seeing them everywhere (which they are) and become paranoid and jaded. Paranoia is healthy for a security professional, but it shouldn't be crippling. People will attempt to influence you through peripheral means, but this is normal, and it doesn't mean that there is *not* a central argument that you should accept for what they are saying.

2.1 *Perceived self-interest*

We will shortly start talking about different kinds of peripheral persuasion strategies, as these are extremely useful for a social engineer to understand, and for others to understand when analysing social engineering. However, before we do, it is worth stressing that the central persuasion route is at least as important, and perhaps more so. A person is far more likely to do what you

[9] It is also possible to *use* peripheral persuasion strategies without having any conscious awareness of doing so. We are social creatures, and we have hardwired instincts for this.

want them to do if they believe it would be in their own interests to do so.

In many cases, what a cybercriminal will be persuading someone to do will not be in anyone's interests but their own. The solution to this contradiction is, of course, to lie. Central persuasion is not somehow a direct route to truth — people can only consider the case put to them. There might be no real reason for me to leave my office unlocked and go outside, but if you tell me the building is about to explode, I can still perceive it to be in my interest to rush off, so long as I believe you.

Much of social engineering therefore rests on selling someone on a lie — something often accomplished through peripheral persuasion strategies. But you should also pay attention to how the truth can be weaponised. Crime can be rational, and a poorly-paid, over-worked employee could certainly personally benefit from ignoring some part of the security policy they are meant to follow[10].

The structure we'll be adopting to look at peripheral persuasion strategies is that of Robert Cialdini, a social psychologist. He studied a great number of "compliance practitioners" — professional persuaders, mostly people like salesmen or politicians, who try to sell something to the public as part of their job — in order to extract six key principles that describe the varied tactics they use. Here's what he had to say about the fundamental nature of perceived self-interest:

> "It is worthy of note that I have not included among the six principles the simple rule of material self-interest—that people want to get the most and pay the least for their choices. This omission does not stem from any perception on my part that the desire to maximize benefits and minimize costs is unimportant in driving our decisions. Nor does it come from any evidence I have that compliance professionals ignore the power of this rule. Quite the opposite: In my investigations, I frequently saw practitioners use (sometimes honestly, sometimes not) the compelling "I can give you a good deal" approach. I choose not to treat the material self- interest rule separately in this book because I see it as a motivational given, as a goes-without-saying factor that deserves acknowledgment but not extensive description."
>
> — Cialdini [13, p. vii]

2.2 *Reciprocity*

The reciprocity principle states that I can convince you to do something for me if I first do something for you. Or, more concretely, I can create a sense of *indebtedness* in you by giving you something, and then exploit your urge to settle that debt. An important point: this isn't just trading behaviour, which

[10] Later in the course we'll talk about security economics, which looks at how we tackle security by looking at these incentive structures.

can be quite rational. The leverage here is that the value you place on what I do for you does not necessarily have to match the value I extract from you. So I can give you something that costs me little, which you might not even want, and use that to pressure you so I get a big reward.

You may have noticed some examples of this around you. For example, charities sometimes send out envelopes containing little token gifts like pencils, notebooks or cards. The reason they do this is because it measurably increases the chance of a donation in return [14]. This isn't rational — the charity's cause isn't any more worthy because they gave you a pencil — but it works. Some charitable organisations take it even further. The Hare Krishna Society was famous for soliciting donations by handing out flowers *and refusing to take them back* – creating a sense of obligation that could not be discharged without giving them money [13, p. 18].

The obligation to repay debts — not just monetary, but social — is powerful, and possibly a human universal[11]. Few people anywhere are happy to receive a gift from someone and then immediately refuse a request from the gift-giver. It is, as a result, a vulnerability which is pervasively exploited. Politicians can be made to vote for policies they don't like because they owe a favour to someone, corporations can benefit a surprising amount by giving out 'free samples' of their products, and criminals get past secretaries because they gave them a box of chocolates once.

There are important aspects to the power of reciprocity. For example, it applies to gifts you did not ask for, and there are usually strong pressures in favour of accepting gifts even if you do not really want them. This means someone can give you a gift (which you don't want) and then get a favour out of you in return (which you don't want to do), which might be of more value than the gift you never wanted.

So, beware strangers bearing gifts. But reciprocity does not stop at the more obvious forms of gift or favour. Reciprocity can also appear as a form of *concession*, and in this form it can be a lot more subtle and deceptive.

If you've ever encountered a window salesman or other high-pressure sales professional, you might have encountered this. The salesman does his measurements, waves a catalogue at you and comes up with a quote for, let's say, £7,000. Your face falls, you might even get angry – that's a crazy price, there's no way you'd agree to that! The salesman is sympathetic, but those are the numbers he has for the materials and labour. He goes over his calculations again, and then wonders... hmm... he wouldn't usually do this but maybe there's a way to get a price reduction. He'll have to ring his manager or supplier. You see one side of a phone call where he appears to be arguing quite vociferously on your behalf. Finally, he turns to you, a bit drained. He can get you a quote for just £4,000, if you agree now.

Of course, everything you think you have just witnessed was a sham. The

[11] A 'human universal' is a rule of human behaviour which is observed in all known cultures, rather than e.g., only in Western society. Psychology has historically been known to confuse the two.

salesman had no real expectation that you would take the first price he offered – the real price was the one he's come down to. And more, you get the distinct feeling of an obligation – he's done so much to bring the price down, it would be cruel of you to reject it. Even if, really, you're not happy with the second price either, you feel like you *owe him* the sale. This is the reciprocity principle in action. The salesmen has 'given' you a substantial reduction from his initial ridiculous price. This tactic of seeking a rejection in order to better reach your real goal is seen throughout sales practice. Here's an example from Cialdini:

> "I was walking down the street when I was approached by an eleven- or twelve-year-old boy. He introduced himself and said that he was selling tickets to the annual Boy Scouts circus to be held on the upcoming Saturday night. He asked if I wished to buy any at five dollars apiece. Since one of the last places I wanted to spend Saturday evening was with the Boy Scouts, I declined. "Well," he said, "if you don't want to buy any tickets, how about buying some of our big chocolate bars? They're only a dollar each." I bought a couple and, right away, realized that something noteworthy had happened. I knew that to be the case because: (a) I do not like chocolate bars; (b) I do like dollars; (c) I was standing there with two of his chocolate bars; and (d) he was walking away with two of my dollars." — [13, p. 27]

The question arises as to whether the boy even had any tickets to sell. This technique appears outside of sales, too — many political debates begin with outrageous demands from one side, with the aim of getting support for their real, more moderate suggestions once those demands are rejected. Calls to "abolish the police", for example, might make people more amenable to discussing the police reform you really want[12]. This concession-based exploit of the reciprocity principle is often known as the 'door-in-the-face' technique, and it is known to have powerful effects on sales [15], and a range of other applications [16, 17, 18].

The concession-based form of reciprocity is ideal for social engineering because you can gain the powerful benefits of a reciprocation obligation without actually having to hand anything over. You can give a sales pitch to an engineering team, knowing that they will turn down your ridiculously overpriced software, and then ask to see their server room on your way out and get the technical specifications you need to craft your malware. You can demand to see the CEO, knowing that when that fails, you are more likely to be able to extract his email address from the stern receptionist. Planning to retreat from an initial approach can be very powerful.

[12]Though there are limits — if your initial demands are *too* extreme, you risk being dismissed as ridiculous and not worth talking to at all.

2.3 *Commitment & Consistency*

The consistency principle states that I can persuade you to a course of action by leaning on your inherent desire to view yourself as consistent. This desire is in itself of course perfectly sensible – nobody wants to deal with an inconsistent person who doesn't keep their word, and so nobody wants to *be* such a person either, however much it might sometimes be in their favour.

Psychologists have long known that commitments can be very powerful for altering behaviour. A favoured example is the 'responsive bystander' result: in a public location where someone has temporarily left their things alone, bystanders are significantly more likely to intervene to stop a staged theft if they have previously agreed to a simple request to "watch my things" [19, 20].

"Interesting," you might comment, "but what's the problem? After all, people don't tend to commit themselves to things they don't want to do."

The exploit in the consistency principle is that individual small, reasonable commitments can be compounded over time into large behaviour changes. Or more simply, if I can get you to first agree to a small request, I can later get you to agree to a much larger request, so long as that request is consistent with the first. This is known as the 'foot-in-the-door' technique.

A good example comes from one of the first experiments: a researcher posing as a volunteer worker went door-to-door in a neighbourhood asking people to allow a (large, ugly) public service billboard about road safety to be erected on their lawn. Most people (83%) refused. However, one particular group mostly (76%) agreed. The difference between this group and their neighbours was that two weeks ago these people had been asked to host much smaller 'be a safe driver' signs [21]. This group had, due to the trivial previous commitment, started to view themselves as someone who supports road safety — and to be consistent with that view of themselves, they now were willing to agree to the much larger billboard. This phenomenon has been replicated in several studies (not always with the same strength of effect), including in online communication [22]. Getting people to agree to a small request improves your odds of getting them to agree to a much less reasonable later request.

This pressure, appropriately and patiently applied, can cause dramatic shifts. Consider this example from Chinese prison camps in the Korean War:

> "An examination of the Chinese prison-camp program shows that its personnel relied heavily on commitment and consistency pressures to gain the desired compliance from prisoners. Of course, the first problem facing the Chinese was how to get any collaboration at all from the Americans. These were men who were trained to provide nothing but name, rank, and serial number. Short of physical brutalization, how could the captors hope to get such men to give military information, turn in fellow prisoners, or publicly denounce their country? The Chinese answer was elementary: Start small and build.

> For instance, prisoners were frequently asked to make statements so mildly anti-American or pro-Communist as to seem inconsequential ("The United States is not perfect." "In a Communist country, unemployment is not a problem."). But once these minor requests were complied with, the men found themselves pushed to submit to related yet more substantive requests. A man who had just agreed with his Chinese interrogator that the United States is not perfect might then be asked to indicate some of the ways in which he thought this was the case. Once he had so explained himself, he might be asked to make a list of these "problems with America" and to sign his name to it. Later he might be asked to read his list in a discussion group with other prisoners. "After all, it's what you really believe, isn't it?" Still later he might be asked to write an essay expanding on his list and discussing these problems in greater detail.
>
> The Chinese might then use his name and his essay in an anti-American radio broadcast beamed not only to the entire camp, but to other POW camps in North Korea, as well as to American forces in South Korea. Suddenly he would find himself a "collaborator," having given aid to the enemy. Aware that he had written the essay without any strong threats or coercion, many times a man would change his image of himself to be consistent with the deed and with the new "collaborator" label, often resulting in even more extensive acts of collaboration." — [13, p. 53]

2.4 *Social Proof*

Why, in that example, were the Chinese having captive soldiers repeat their list of statements to other soldiers? Part of the answer is to do with the commitment strategy – a public statement is a larger commitment, and so will move the speaker more towards the stance you want them to adopt. The other part relates to the principle of *social proof*, and the psychological effect of hearing other American POWs criticise America.

The principle of social proof states that one means we rely on to determine what is correct is the opinions of other people. In particular, large groups of people, or people 'like us' in some way. This can be sensible political behaviour – sometimes even if everyone else is wrong it might be safer for you to go along with them than disagree – but has obvious flaws when it comes to truth-seeking, which is why we try to agree on norms of open discussion, evidence-checking and debate that limit the tendency towards 'groupthink'.

The major security issue with social proof is that it can be fabricated. This was true even before the internet – many con-men make use of planted collaborators in their audience to control the behaviour of the crowd and manage seemingly 'spontaneous' group behaviour like offering up donations. More broadly, anyone with influence over a medium (like a newspaper editor,

or their funder) can create the impression of widespread agreement with a position[13]. With the web now allowing for disconnection between real individuals and their apparent presence, social proof is even easier for people and businesses to manufacture. Just because a website *tells you* that 3 other people are looking at the holiday package you're considering doesn't mean those people really exist. What look like reviews from satisfied customers might be shills – paid-for positive feedback from automated services.

For social engineers attempting to gain access to systems, social proof is most important for establishing a *pretext* – an invented identity that has a valid reason for making the requests they initiate. The exact forms can be varied: a falsified email chain purportedly with someone else in the organisation approving the request; a Twitter account that seems to show this is a real person with a reasonable number of followers; a malicious executable hosted on a page which also presents 'comments' from seemingly satisfied and safe users of the software.

2.5 *Liking*

We're biased towards people we like. That sounds so obvious as to be almost tautological. Yet the expression of this bias is often subtle, and can operate below the level of our awareness. Quite often people do not realise that they are treating others differently based on how they feel about them, and might reject the idea even when it is demonstrated empirically. As a social engineer, getting people to like you can make a lot of things easier, including getting them to believe something they shouldn't, and making them less likely to suspect you.

There are lots of reasons individuals can like each other, not all of which are easily described or manipulated. One workaround that marketing professionals have used is to use *pre-existing* friendships to sell their brand. Customer-led marketing of products like Tupperware and Avon (and nowadays various 'essential oils' products and multi-level marketing scams) gets members of the public to sell the product to their friends, who may feel less comfortable rejecting them than they would a salesperson. Social engineers also use this vulnerability, using compromised accounts to send phishing emails or other fraud to the account's contacts, who are more likely to respond to a request from somebody they think they know and trust.

For a social engineer unable to leverage such pre-existing connections (or invest sufficient time to develop them), there are a couple of key attributes that affect instant likeability. First of all, *physical attractiveness* can be incredibly powerful for affecting how people treat you. To take some examples from studies of judges' sentencing decisions, physically unattractive people received fines up to 300% larger than those for attractive people [24], and got prison sentences between 51% [25] and 119% [26] longer (the former figure from

[13] If you're interested in the question of how this can be applied at a societal level, see Herman & Chomsky's book 'Manufacturing Consent' [23].

mock trials, the latter figure from actual sentences). At work, even for jobs that have nothing to do with looks, employers typically want to pay beautiful workers higher wages [27]. You may already have heard of the 'halo effect' – if you have one good quality, people tend to perceive you as better in other respects. Physical attractiveness can be an opening to being considered honest, reliable and trustworthy – all incredibly useful for a social engineer.

Another important factor is *similarity* – we like people who appear to be similar to us, whether that similarity is in opinions, personality traits, style of clothing, hobbies, or verbal mannerisms. For just this reason, salespeople have long been trained to identify the interests and style of customers and imitate them [13, p. 131]. If you find yourself liking someone who's trying to sell you something, you should be very suspicious. These same tactics can be used by hustlers of the criminal variety.

There are many more strategies for getting into people's good graces, from straightforward compliments to the sort of operant conditioning involved in giving someone a reward (a smile, a sweet, a hug) whenever they see you. In many cases, the mechanism is secondary to the degree of subtlety involved in its application and, from the social engineer's perspective, the amount of time investment necessary before it pays off in terms of the access being sought. Cheap ways to make yourself more likeable are therefore valued. For example, an important low-cost tactic for many online scammers is pretending that they are involved in some worthy cause, like working for a charity, in order to buy themselves some instant goodwill.

2.6 *Authority*

The authority principle states that people are much more likely to follow instructions from others in positions of authority over them. This has an obvious logic — few organisations would survive if nobody did what their superiors told them to do — but this tendency can be much more far-reaching, including people being willing to ignore not only official policy when following orders, but their own moral instincts and common sense.

For example, in a worrying study, researchers found that 21 of 22 nurses would administer a clear overdose of an unfamiliar drug to a patient when ordered to by a person who identified himself on the telephone as a doctor [28]. This was despite explicit hospital policy against taking instructions over the phone, the drug not being approved for administration, the doctor not being familiar to them, and none of the required paperwork being filled out[14]. Even more chillingly, in at least 70 real criminal acts between 1992 and 2004, an anonymous caller would ring fast food restaurants such as McDonald's, pretending to be a law enforcement officer investigating a crime. Using this invented authority, the caller would get the manager of the store to detain and

[14] An attempted replication of this study found that nurses were much more likely to question the order if they were more familiar with the drug and able to talk to other nurses about the order [29].

strip-search employees or customers they accused of petty larceny[15], often escalating to orders for sexual assault [30].

One of the most famous studies on obedience to authority was that of Stanley Milgram [31]. In this experiment, participants played the role of an assistant in a study on memory, administering progressively stronger electric shocks to an actor pretending to be the experimental subject. The real object of the study was whether the participants would refuse to continue to shock the subject (who acted like they were in pain) when prodded to continue by the seeming authority – the man in the lab coat. The researchers expected that most people would stop around a point marked 'Very Strong Shock'. The results were that 65% of people continued to the highest possible shock level (two levels beyond the point marked 'Danger: Severe Shock' on the shock generator they were operating). While some of Milgram's method, and the interpretation of his results has been questioned [32], the results have been replicated in a number of followup studies [33, 34].

The benefit of the authority principle for a social engineer is that authority can be faked, sometimes shockingly easily. As demonstrated in the nursing study and the strip-search calls, simply presenting yourself as an authority figure can sometimes be enough. Indicators of authority such as the way you are dressed, or the number of impressive titles in your email signature can help solidify such claims. Stolen or borrowed authority can be even better. An entire class of fraud known as 'business email compromise' revolves around capturing or impersonating a senior figure's email account in order to make demands of employees (such as suddenly changing the bank account to which funds are transferred). When put to the test, many employees will follow the instruction their boss apparently just gave them rather than the policy they were originally trained to follow. Even where impersonation is not possible, referencing other authorities is a powerful pressure that also shields the social engineer from questions — if the person on the phone says they were instructed to do something by the company's CEO, then refusing to help them suddenly becomes a riskier prospect for the helpdesk employee or IT admin, and questions like "why" or "what about…" can be deflected up the hierarchy, where they are unlikely to actually be asked.

2.7 *Scarcity*

We're running out of principles. You should grab this one quickly, before it's gone! The scarcity principle is that opportunities seem more valuable when their availability is limited. By stressing that an option is available only for a limited time or for a select few people, you can greatly enhance its desirability. When you're considering buying something, salespeople will often tell you that there are very few available, for precisely this reason. In fact, Cialdini reports that sometimes they might tell you there are *no* such items available:

[15] Such searches are themselves illegal.

> The tactic was played to perfection in one appliance store I investigated, where 30 to 50 percent of the stock was regularly listed as on sale. Suppose a couple in the store seemed from a distance to be moderately interested in a certain sale item. There are all sorts of cues that tip off such interest—closer-than-normal examination of the appliance, a casual look at any instruction booklets associated with the appliance, discussions held in front of the appliance, but no attempt to seek out a salesperson for further information. After observing the couple so engaged, a salesperson might approach and say, "I see you're interested in this model here, and I can understand why; it's a great machine at a great price. But, unfortunately, I sold it to another couple not more than twenty minutes ago. And, if I'm not mistaken, it was the last one we had."
>
> The customers' disappointment registers unmistakably. Because of its lost availability, the appliance jumps suddenly in attractiveness. Typically, one of the customers asks if there is any chance that an unsold model still exists in the store's back room, warehouse, or other location. "Well," the salesperson allows, "that is possible, and I'd be willing to check. But do I understand that this is the model you want and if I can get it for you at this price, you'll take it?" Therein lies the beauty of the technique. In accord with the scarcity principle, the customers are asked to commit to buying the appliance when it looks least available—and therefore most desirable. Many customers do agree to a purchase at this singularly vulnerable time. Thus, when the salesperson (invariably) returns with the news that an additional supply of the appliance has been found, it is also with a pen and sales contract in hand. The information that the desired model is in good supply may actually make some customers find it less attractive again. But by then, the business transaction has progressed too far for most people to renege. The purchase decision made and committed to publicly at an earlier, crucial point still holds. They buy. — [13, p. 180–181]

Note the interplay of principles here—false scarcity is used to elicit a commitment, which then cannot be reneged upon. You might also detect an element of reciprocity in the script, with the salesman stressing the personal effort they're putting in to the transaction, as something that looks a little like a favour. Successful influence of behaviour rarely relies on just one principle.

Another major application of scarcity is in time pressure, where you say that you need a decision quickly. This can be especially powerful for reasons beyond the attractiveness of limited availability. The 'cognitive', central route to decision-making is necessarily slower, as it takes time to consciously assess the case being put to you. It often takes time and reflection to detect an inconsistency or lie[16]. People not only make poorer decisions generally when

[16]Some of you might notice that you tend to recognise plot holes in TV shows *after* you've

you put them under time pressure, the training they might have had for this situation is far less effective [35].

Social engineers can impose time pressure by creating deadlines — "This needs to be done by 4pm" which panic people so they spend little time evaluating whether the demand is legitimate. However, they can also exploit existing time pressures. An overworked employee will have little time to think about applying security policies, and somebody just reaching the end of a working week will be clearing incoming tasks with as little effort as possible[17].

3 OPTIONAL EXERCISES

These are suggestions for students who are particularly interested in the topic of social engineering and the theory underlying behavioural influence. They are not assessed in any way. You don't need to complete the below to do well on the course.

3.1 *Casual Viewing*

Social engineering turns up a lot in fiction, from the cult movie *Hackers* to entire series like *White Collar*. Often these fictional versions reflect some real social engineering mechanisms, but they may be exaggerated or simplified. Even when a production is based on true events, such as in *Catch Me If You Can*, reflecting Frank Abagnale's life, or *Compliance*, a movie based on the strip-search scam calls, the fact that everything is scripted means you cannot really rely on what you see. Social engineering is an art that requires honed skill to carry off against real people, rather than intentionally-gullible characters. One of the best options is a BBC show *The Real Hustle*, which demonstrates a number of in-person cons on real members of the public[18]. The format has since been copied by a few international broadcasters.

3.2 *Further Reading*

A number of books and papers are cited in these notes, with full references given below. Many of these would be worth reading, depending on which elements you are most interested in. The two books with the most general relevance to these lectures are Mitnick & Simon's *The Art of Deception* [4], and Cialdini's *Influence: The Psychology of Persuasion* [13].

Mitnick & Simon present a number of 'example' social engineering scripts with analysis—given that Mitnick is an infamous social engineer, and other successful criminals like Kane Gamble have used these techniques, this is

finished watching them, when you've had time to realise that something the show did doesn't make sense, instead of while you're having to pay attention to what's going on on-screen.

[17] A lot of major cybersecurity attacks happen on Friday afternoons, in part because of this lapsed attention from employees, and in part because the weekend typically presents the prospect of two days to secure your hack and hide your traces

[18] Though, true to style for a cast of conmen, some scenes turned out to in fact be staged.

clearly valuable insight. More than a few of his 'examples' are suspected of being slightly disguised versions of real cons he has carried out in the past.

I have presented a few quotes from Cialdini's book in these notes, which I hope convince you that he is an entertaining writer. Despite presenting a lot of academic psychology material, *Influence* is a light and fun read, filled with good examples and even practical advice for resisting each of the persuasion principles. If you wanted a longer, better version of these lecture notes, this is what I would recommend.

3.3 *In Practice*

Opportunities to safely practice deceiving people are rare in everyday life[19]. Fortunately, there are some ways these skills can be honed. For online, predominantly text-based deception, social engineering often boils down to a creative writing task. Practicing the fundamentals of good writing, including the deliberate control of your writing style, is something you can do either alone or as part of a group (online via writing forums or offline via workshops and creative writing programmes). A stint as a journalist (say, on a student paper) would also give you practice finding people and extracting information from them.

For practicing in-person or voice-based social engineering, the best approach is probably taking *acting* classes. Acting is after all entirely about pretending to be someone else, including all the subtle mannerisms, stances and conversation styles. Local theatre groups are a good place to start, and particularly quick-fire *improv* acting sessions, which force you to think on your feet in an evolving, unscripted conversation—just like a real social engineer must.

REFERENCES

References

[1] R. Munroe, "XKCD 538: Security," [Online] https://xkcd.com/538/, accessed 2020-06-03.

[2] E. F. Vidocq, *Memoirs of Vidocq: Principal Agent of the French Police Until 1827*. Carey and Hart, 1844.

[3] S. Padgett, *Profiling the Fraudster: Removing the Mask to Prevent and Detect Fraud*. John Wiley & Sons, 2014.

[4] K. D. Mitnick and W. L. Simon, *The Art of Deception: Controlling the Human Element of Security*. John Wiley & Sons, 2003.

[5] K. Hafner and J. Markoff, *Cyberpunk: Outlaws and Hackers on the Computer Frontier*. Simon & Schuster, 1991.

[19] Even when you're not breaking any law, most people will eventually get annoyed with you.

[6] K. Hafner, "Kevin Mitnick, unplugged." *Esquire*, vol. 124, no. 2, pp. 80–88, 1995.

[7] N. Massey, "Kane Gamble: Teenager with autism on Leicestershire housing estate took classified information by fooling people into thinking he was FBI boss," *The Independent*, accessed 2020-06-05. [Online]. Available: https://www.independent.co.uk/news/uk/crime/us-intelligence-cia-fbi-american-government-john-brennan-mark-giuliano-crackas-with-attitude-latest-a8170561.html

[8] H. Dixon, "British 15-year-old gained access to intelligence operations in Afghanistan and Iran by pretending to be head of CIA, court hears," *The Telegraph*, accessed 2020-06-08. [Online]. Available: https://www.telegraph.co.uk/news/2018/01/19/british-15-year-old-gained-access-intelligence-operations-afghanistan/

[9] Verizon Enterprise Solutions, "Data breach investigations report (DBIR)," Tech. Rep., 2020, accessed 2020-06-08. [Online]. Available: https://enterprise.verizon.com/resources/reports/dbir/

[10] Anti-Phishing Working Group (APWG), "Phishing attack trends report – 1st quarter 2020," Tech. Rep., 2020, accessed 2020-06-08. [Online]. Available: https://apwg.org/trendsreports/

[11] ——, "Phishing attack trends report – 4th quarter 2020," Tech. Rep., 2020, accessed 2021-09-07. [Online]. Available: https://apwg.org/trendsreports/

[12] R. E. Petty and J. T. Cacioppo, "The elaboration likelihood model of persuasion," in *Communication and Persuasion*. Springer, 1986, pp. 1–24.

[13] R. B. Cialdini, *Influence: The psychology of persuasion*. William Morrow & Company, 1993.

[14] A. Falk, "Gift exchange in the field," *Econometrica*, vol. 75, no. 5, pp. 1501–1511, 2007.

[15] C. Ebster and B. Neumayr, "Applying the door-in-the-face compliance technique to retailing," *The International Review of Retail, Distribution and Consumer Research*, vol. 18, no. 1, pp. 121–128, 2008.

[16] A. C.-y. Chan and T. K.-f. Au, "Getting children to do more academic work: Foot-in-the-door versus door-in-the-face," *Teaching and Teacher Education*, vol. 27, no. 6, pp. 982–985, 2011.

[17] P. W. Eastwick and W. L. Gardner, "Is it a game? Evidence for social influence in the virtual world," *Social Influence*, vol. 4, no. 1, pp. 18–32, 2009.

[18] A. Pascual and N. Guéguen, "Door-in-the-face technique and monetary solicitation: An evaluation in a field setting," *Perceptual and Motor Skills*, vol. 103, no. 3, pp. 974–978, 2006.

[19] T. Moriarty, "Crime, commitment, and the responsive bystander: Two field experiments." *Journal of Personality and Social Psychology*, vol. 31, no. 2, p. 370, 1975.

[20] N. Guéguen, M. Dupré, P. Georget, and C. Sénémeaud, "Commitment, crime, and the responsive bystander: effect of the commitment form and conformism," *Psychology, Crime & Law*, vol. 21, no. 1, pp. 1–8, 2015.

[21] J. L. Freedman and S. C. Fraser, "Compliance without pressure: the foot-in-the-door technique." *Journal of Personality and Social Psychology*, vol. 4, no. 2, p. 195, 1966.

[22] N. Guéguen, "Foot-in-the-door technique and computer-mediated communication," *Computers in Human Behavior*, vol. 18, no. 1, pp. 11–15, 2002.

[23] E. S. Herman and N. Chomsky, *Manufacturing consent: The political economy of the mass media*. Pantheon Books, 2002.

[24] A. C. Downs and P. M. Lyons, "Natural observations of the links between attractiveness and initial legal judgments," *Personality and Social Psychology Bulletin*, vol. 17, no. 5, pp. 541–547, 1991.

[25] A. Desantts and W. A. Kayson, "Defendants' characteristics of attractiveness, race, and sex and sentencing decisions," *Psychological Reports*, vol. 81, no. 2, pp. 679–683, 1997.

[26] J. E. Stewart, "Defendant's attractiveness as a factor in the outcome of criminal trials: An observational study 1," *Journal of Applied Social Psychology*, vol. 10, no. 4, pp. 348–361, 1980.

[27] M. M. Mobius and T. S. Rosenblat, "Why beauty matters," *American Economic Review*, vol. 96, no. 1, pp. 222–235, 2006.

[28] C. K. Hofling, E. Brotzman, S. Dalrymple, N. Graves, and C. M. Pierce, "An experimental study in nurse-physician relationships," *The Journal of nervous and mental disease*, vol. 143, no. 2, pp. 171–180, 1966.

[29] S. G. Rank and C. K. Jacobson, "Hospital nurses' compliance with medication overdose orders: a failure to replicate," *Journal of Health and Social Behavior*, pp. 188–193, 1977.

[30] A. Wolfson, "A hoax most cruel: Caller coaxed mcdonald's managers into strip-searching a worker," *Louisville Courier-Journal*, p. 142, 2005.

[31] S. Milgram, "Behavioral study of obedience." *The Journal of Abnormal and Social Psychology*, vol. 67, no. 4, p. 371, 1963.

[32] G. Perry, "Deception and illusion in Milgram's accounts of the obedience experiments," *Theoretical & Applied Ethics*, vol. 2, no. 2, pp. 79–92, 2013.

[33] T. Blass, "The Milgram paradigm after 35 years: Some things we now know about obedience to authority 1," *Journal of Applied Social Psychology*, vol. 29, no. 5, pp. 955–978, 1999.

[34] J. M. Burger, "Replicating Milgram: Would people still obey today?" *American Psychologist*, vol. 64, no. 1, p. 1, 2009.

[35] D. Zakay and S. Wooler, "Time pressure, training and decision effectiveness," *Ergonomics*, vol. 27, no. 3, pp. 273–284, 1984.