Every ROSE has its thorn

The dark art of Remote Online Social Engineering

Matt Wixey

Research Lead, Cyber Security

Building a secure digital society.

pwc

www.pwc.com

Introduction

Matt Wixey

- Research Lead for the Cyber Security BU
- Work on the Ethical Hacking team
- PhD student at UCL
- Previously worked in LEA doing technical R&D

Disclaimer

- This content is presented **for educational purposes only**
- What this presentation isn’t…

Preface

“And as imagination bodies forth
The forms of things unknown, the poet's pen
Turns them to shapes and gives to airy nothing
A local habitation and a name.”

William Shakespeare, *A Midsummer Night's Dream,* 5.1, 14.

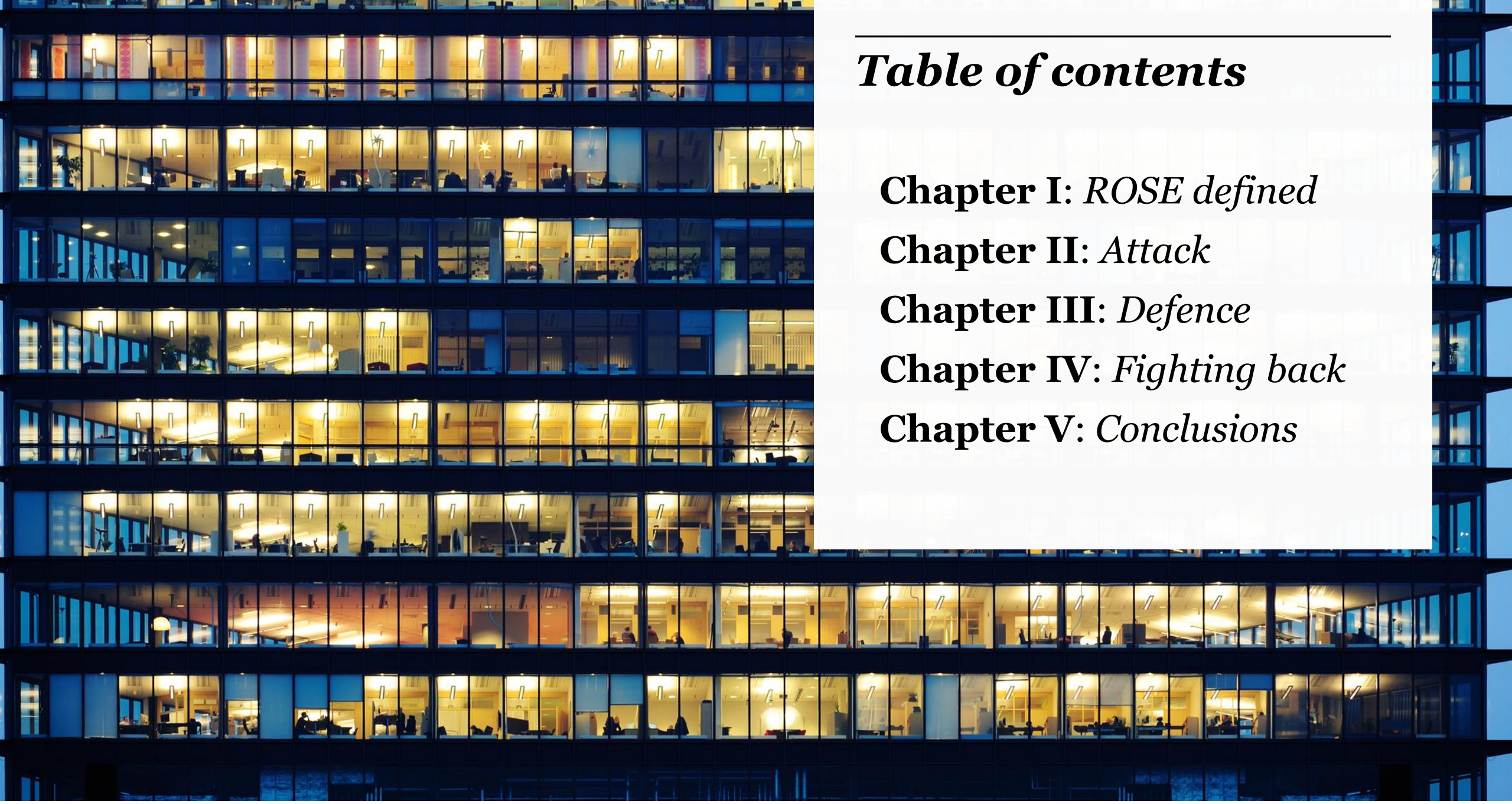


Table of contents

Chapter I: ROSE defined

"We are never deceived; we deceive ourselves."

Johann Wolfgang von Goethe, *Sprüche in Prosa*, III.

Traditional online deception types

Trolling

- *The art of trolling* – Matt Joyce, DEF CON 19
- Sophistry & fallacies to provoke responses
- Often used as shorthand for any online abuse

Sockpuppetry

- Often short-term, light on detail
- Posed as independent
- Operated by same entity
- Stealth marketing, false reviews, inflating polls

Astroturfing

- Sub-category of sockpuppetry
- Used to influence policy, manipulate consensus
  - Especially in politics and marketing
- *Julius Caesar*

Phishing

- Mass phishing
- Spear-phishing
- Whale-phishing
- Interesting taxonomy
- Proposing some new additions…

New phishing categories

Octo-phishing

- Targeting 8 people at a time

Crab-phishing

- 2 attackers phish a target
- So it’s a pincer movement

New phishing categories

Loch Ness monster-phishing

- When you're not sure your targets even exist…
- …but you once saw a grainy black-and-white photo of them

Dead Sea-phishing

- When your targets don't even have internet

New phishing categories

Kraken-phishing

- Incredibly high-risk attack
- You phish one huge, dangerous, mythical target in an epic but ultimately doomed attempt
- Your entire infrastructure is completely destroyed
- The vortex claims your folly

Catfishing

- Long-term false personae; one-on-one interaction
- Targeting: specific or randomly selected
- Motivations (not exhaustive)
  - Psychological e.g. attention-seeking (Magdy et al, 2017)
  - Fraud, extortion

ROSE

- Long-term, self-referenced, highly customised
- Maintained manually and highly interactive
- One or more detailed false persona
- **Focus on business-related platforms and targets**
- **Objective of compromising security**

Security and fiction writers

- ROSE *is* creating fiction
  - Believable characters and backstories
  - Realistic dialogue, reactions, compelling plot
  - Realistic conclusions

Case study I: Mark and John (UK, 2003)

- Knew each other in real life
- John: multiple personae on MSN, targeting Mark
- Complex and sophisticated, with corroboration
- Unclear motivation – sexual elements

Case study I: Mark and John (UK, 2003)

- One persona: high-ranking Mi5 officer
  - Persuaded Mark to kill John
  - Mark stabbed John, called him an ambulance
- Mark convicted of attempted murder
- John convicted of inciting *his own* murder

https://www.vanityfair.com/news/2005/02/bachrach200502

Case study II: Robin Sage (US, 2010)

- "Cyber threat analyst", 10 years' experience
  - 25 years old!
- 28 days, 300 connections/friends
- Multiple invitations, offers
- No warning signals shared!

https://www.privacywonk.net/download/BlackHat-USA-2010-Ryan-Getting-In-Bed-With-Robin-Sage-v1.0.pdf

Case study III: Abby Pierce (US, 2010)

- Photographer befriends Abby (8 y.o.) and family
- Becomes romantically involved with older sister
- Abby's mother, Angela, was the older sister
  - Maintained ~15 online personae
  - All communicated with each other

Chapter II: Attack

"O, what a tangled web we weave,
When first we practice to deceive!"

Sir Walter Scott, *Marmion,* V1.17

"But when we've practiced quite a while,
How vastly we improve our style!"

J.R. Pope, *A Word of Encouragement*

Why do it?

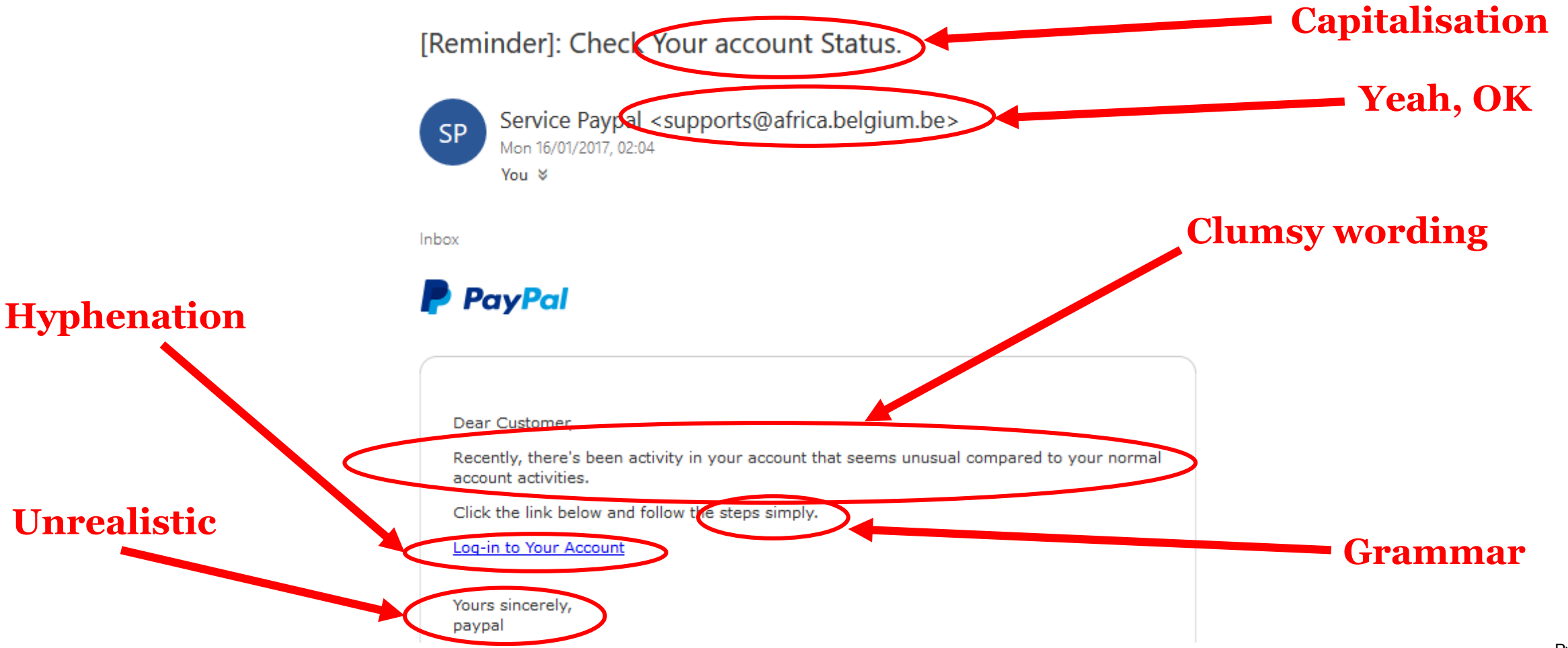
- To bypass controls and effects of user education
- Multiple attack opportunities
- To bypass filters…

Filters

- We all have a set of filters and thresholds
  - Upbringing, education, experience, training, personality
  - Distinctive and consistent (CAPS)
    - Mischel & Shoda, 1995; Michel, 1999; Zayas et al, 2002; Shoda et al, 1994

Filters

- Mass phishing doesn't really try to bypass filters
  - Other than in a crude sense e.g. self-selection



Filters

- ROSE can be designed, specifically, to bypass *your* filters
- *"There exists, for everyone, a sentence - a series of words - that has the power to destroy you."*
  - Philip K. Dick, *VALIS*

Filters

Realistic approach, relevant to my interests

Real reference

Interest

Referenced in talks

Hi Matt,

Hope you don't mind the unsolicited email! Just wanted to get in touch and say that I saw your DEF CON talk from last year and thought it was great. I particularly enjoyed the part on exfiltrating data by altering screen brightness levels. Shades of Mordechai Guri (and William Gibson!)

To introduce myself, I work for Random Technology Publishing, and wondered if you would be interested in writing a chapter for an upcoming e-book on new attack techniques? We'd be especially interested in unconventional approaches and attack vectors.

Please let me know if you'd be interested in the first instance, and we can discuss remuneration following that?

Happy to discuss further by email or phone (details in signature).

All the best,

I use this

Invites phonecall

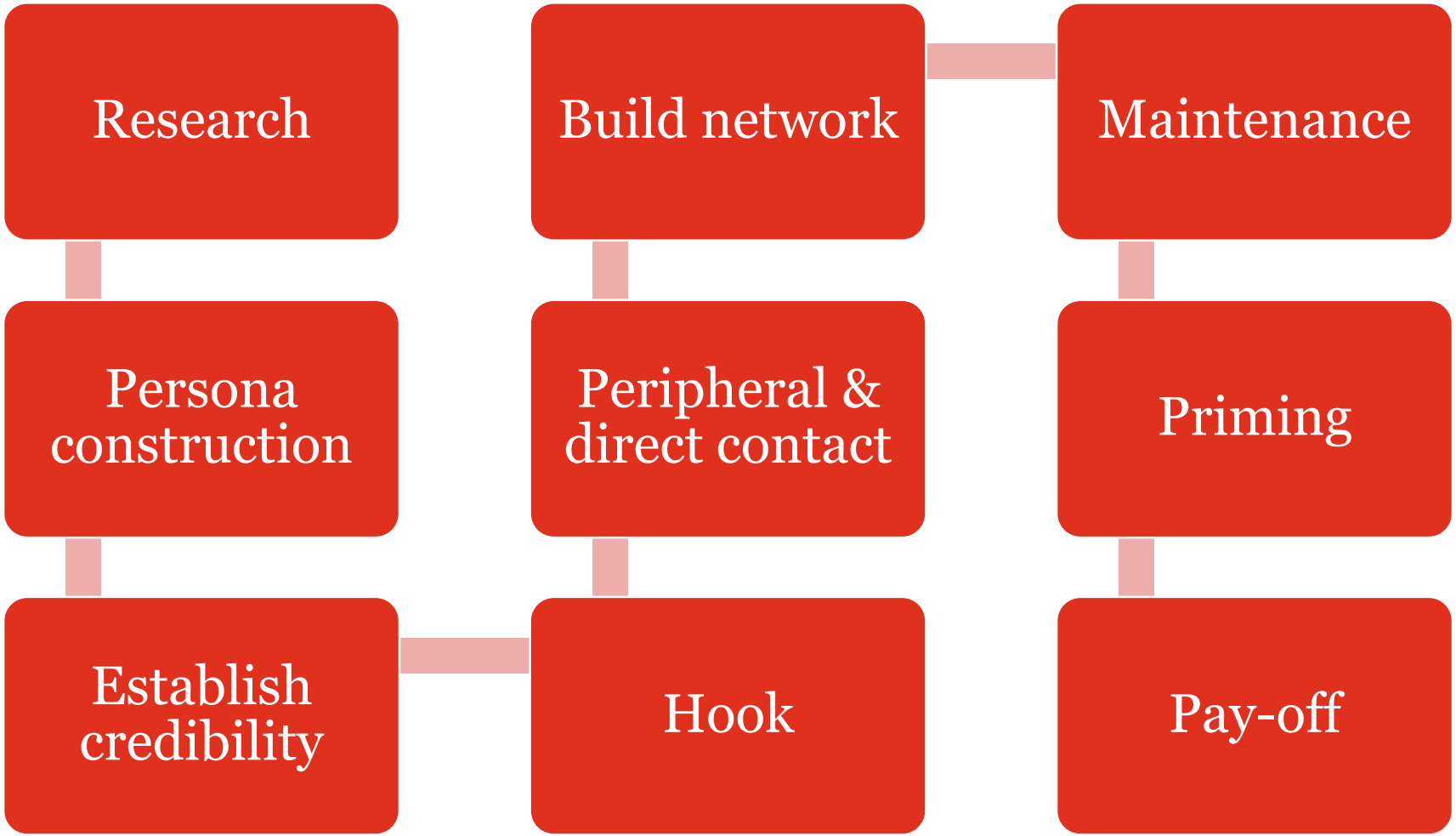
No mention of money initially

John Whitson
New Content Manager
Random Technology Publishing

w: randomtechnology.com
e: john.whitson@randomtechnology.com
t: 020 7946 0000
fb: facebook.com/pages/randomtechnologypublishing
tw: twitter.com/rtpub

These would all be back-stopped, with interactions & history, and even genuine e-books

Methodology outline



Research

Attack

- Specific attributes
- Likes/dislikes, interests, hobbies
- Affiliations
- Education/employment
- Relationships and family
- Locations
- Other platforms and profiles
- Purchases, holidays
- Technical info
- Reactions, style, motivations

Defence

- Limit sensitive information
- Google alerts
- Various services to alert when you've been searched for

Persona construction

Attack

- Mirroring or supplementing target
- Similar interests, styles, etc
- Potential openings for contact
- Profile images
  - Not always stolen
  - May be edited/manipulated
  - Or behind paywall or from private source
  - Or completely new

Defence

- Limit sensitive information
- Google alerts and similar
- New additions to network
- Reverse image search
- Manipulation detection
  - Glitches
  - Error level analysis
  - Lighting, textures, patterns, blurs
- Perceptual hashing
- Metadata e.g. dates, and context

Establishing credibility

Attack

- Referencing institutions, places, companies, etc
- Backdating – not 100% reliable
- Pre-age accounts: create in advance
  - May auto-post for some time
- Profiles which age over time
  - Change images, styles, politics
- Profiles never used for attacks
  - But their “children” are in 20 years
  - Playing the really long game

Defence

- New accounts are suspect
- Backdating can be examined
- Check for early auto-posting (anti-bot analysis)
- Validation (direct or indirect)
- Genuine knowledge: attribution
- Inconsistencies: opportunity
- **Share findings**

Building a synthetic network

Attack

- Proxies, Tor, burner phones, SIM swapping, etc
- More advanced techniques
  - Deepfakes
  - Voice morphing
  - Google Duplex
- Avoiding profile contamination
- Distinctive voices and styles

Defence

- Forensic linguistics
- Behavioural attribution
- Check for profile contamination
- Inconsistencies
- Cultural indicators
- Metadata

Peripheral -> Direct contact

Attack

- Starting with associates
  - Shows on feed = context later
- ‘Like’ same things
- Trying to get into circle of awareness
- cp. *Donnie Brasco* (Pistone, 1988)
- Liking, commenting, adding
- Prefaced w/ reference to peripheral

Defence

- Corroborate with mutual associates

The hook

Attack

- Informed by earlier research
- Could be request for help/advice
- Or something that will benefit
  - Flirting/sexual
  - Business relationships
  - Ambitions/fantasies
- Shift to corporate email
- Reveals background subtly
  - Drip-feed basis

Defence

- Self-assessment
- Understanding your filters
  - Self-assessment of flaws
- Question motivations and consequences
- Ask how corporate email was found
- Question why they want to shift to corporate email
- Consider ‘sandboxing’ on social media

Maintenance

Attack

- Frequent contact
- Adapted to reality e.g.
  - Local holidays and events
  - Office hours, timezones
  - Appropriate IP and geolocation
- Adapts to responses and context
- Building rapport and trust
- Draws target into synthetic web
  - Use other profiles to communicate
  - Insurance
  - Other angles and opportunities

Defence

- Forensic linguistics
- Behavioural attribution
- Check for evasiveness around voice/video/F2F comms
- Inconsistencies and errors

Priming

Attack

- Microcosm
  - e.g. multiple benign attachments
  - or revealing less valuable info
  - or clicking on links
- Obtains technical feedback
- Conditioning
- Small steps to bigger ones

Defence

- Question motivations when asked to do something
- Qs on technical aspects = red flag

The pay-off

Attack

- Launches attack
  - Attachment
  - Link
  - Ask for information
  - Extortion
  - Seed profile with malware
- May maintain contact
  - To re-use profile in future
  - Now with real-world corroboration
- Or may disappear

Defence

- Sudden disappearance or lack of contact/interests = red flag

Case study IV: Mia Ash (Middle East, 2017)

Mia Ash 500+ connections

Photographer at Mia's Photography

London, Greater London, United Kingdom | Photography

Current Mia's Photography

Previous Loft Studios, Clapham Studios

Education Goldsmiths, University of London

- Dell SecureWorks
- London-based photographer
- Profiles developed for at least a year
- Used when initial phishing campaigns failed
- LinkedIn, FB, WhatsApp, Blogger, DeviantArt

https://www.secureworks.com/~/media/Images/Insights/Resources/Threat%20Analyses/078%20mia%20ash/mia-ash-02.ashx?la=en&hash=62CDAE07A741F6852C960462A5BFC87A1ED61B04

Case study V: Syria, 2015

- FireEye
- Profiles targeting Syrian opposition on Skype
- Sent malware disguised as images
  - .pif -> DarkComet
- Profiles sometimes maintained to get further info
- Evidence of research e.g. dates of birth

https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-behind-the-syria-conflict.pdf

Ethics and legality

- I am not a lawyer
  - Grey area – not illegal in the UK to pretend to be someone else online
  - May breach T&Cs
  - Illegal to impersonate specific professions e.g. police officers
  - Psychological effects after engagement

Chapter III: Defence

"A lie never lives to be old."

Sophocles, *Acrisius,* fragment 59.

Challenges for defenders

- Contact may occur outside working hours, or:
  - On private social media, and/or on private equipment
  - And therefore, rightly, outside of remit
- Many employers *want* staff to engage on social media

Understanding deception

- 90% of people expect others to lie online sometimes
  - Drouin et al, 2016
- People lie in 14% of emails, 27% of F2F, and 37% of calls
  - Hancock, 2007
- On average, people lie twice a day
  - DePaulo et al, 1996

Understanding deception

- Average people are really bad at detecting deception!
- Slightly better than chance (Bond & DePaulo, 2006)
- Worse online (George et al, 2004; Zhou & Sung, 2008)
- Rely on faulty cues (Toma & Hancock, 2012)

Why do we fall for it?

- High early trust levels (Kramer, 1994)
  - Unit grouping (Kramer et al, 1996)
  - Reputations (Barber, 1983; Powell, 1996; Dasgupta, 1988)
  - Belief-confirming cognitive mechanisms (Good, 1988; Taylor & Brown, 1988)
- Trust violations often seen as isolated events (Sitkin & Roth, 1993)
- Truth bias (Levine et al, 1999)
- Uncertainty reduction theory (Berger & Calabrese, 1975)

Interpersonal deception theory

- Buller & Burgoon, 1996
- Deception is a goal-driven activity
- Dynamic process, involving exchanges and adaptation
- Liars have 2 primary goals
  - Achieve deception (persuasive)
  - Avoid detection (protective)

Expanded prominence-interpretation theory

- George et al, 2016, building on Fogg, 2003
- Decisions about lying behaviour depends on:
  - Assessment of prominence and reputation of site
  - Assessment of media
  - Past experiences

Media richness theory

- Daft & Lengel, 1986; Daft et al, 1987
- Inherent properties of medium influence comms
- Rich mediums:
  - Instant mutual feedback, verbal and non-verbal cues
  - Availability of natural language, tailored discussion

Media synchronicity theory

- Dennis et al, 2008
- Temporality of feedback and responses
  - Synchronous: IM
  - Asynchronous: email

Communication accommodation theory

- Giles & Coupland, 1991; Pickering & Garrod, 2004
- For persuasion/approval:
  - Match accent, volume, vocabulary, grammar, gestures
  - Align linguistic representations
  - Linguistic style matching (Niederhoffer & Pennebaker, 2002)

Cognitive load

- Complex neural processes involved in lying (Hu et al, 2012)
  - Suppression of truth
  - Strategic behavioural modifications
  - Psychological and physiological arousal (Pak & Zhou, 2013)
  - Police better at detecting overloaded liars (Porter & ten Brinke, 2010)

Strategies, processes and detection

Strategies

- Withhold
- Vagueness and uncertainty
- Non-immediacy (distancing)
- Burgoon et al, 1996; Zhou et al, 2003

Processes

- Psychological processes experienced by liars
- Processes used to accomplish deception
- Toma & Hancock, 2012

Detection

- Leakage cues, sent unwillingly
- Strategic decisions, willingly transmitted
- Tsikerdekis & Zeadally, 2014

Detecting deception

- There is no "*Pinocchio's nose*" (Vrij, 2008a)
- Deception can change with stakes (DePaulo & Kirkendol, 1989)
- Synchronicity, richness, motivation, stakes
- Experience (Granhag et al, 2004; Hartwig et al, 2004)

Linguistic markers

- **Less delay, more participation** (Zhou & Zhang, 2004; Zhou, 2005)
- **Less corrections and edits** (Zhou & Zhang, 2004)
- **More words** (Zhou et al, 2004; Burgoon et al, 2003; Hancock et al, 2005; Zhou et al, 2004; Hancock et al, 2007 – 28%; Ho et al, 2015 – 30%)
- **More informal and uncertain** (Zhou & Sung, 2011)
- **Non-immediacy** (Zhou & Sung, 2011; Hancock et al, 2005; Zhou, 2005; Zhou & Sung 2008; Newman et al, 2003; Hancock et al, 2007; Toma & Hancock, 2010; Keila & Skillicorn, 2005; Zhou et al, 2004)

Linguistic markers

- **Less diverse and less complex** (Zhou & Sung, 2011; Zhou & Sung, 2008; Zhou et al, 2004)
- **Less sensory, more cognitive** (DePaulo et al, 2003; Johnson & Raye, 1981; Hauch et al, 2015; Ho et al, 2015)
- **Avoid topics they've lied about, emphasise truth** (Toma & Hancock, 2012)
- **Motivated liars avoid causal terms; unmotivated liars increase use of negation** (Hancock et al, 2007)
- **More negative emotion words** (Knapp & Comadena, 1979; Newman et al, 2003; Vrij, 2000; Zhou et al, 2004)

Linguistic markers

- **Less exclusive terms and negations**, which mark commitment and specificity (Newman et al, 2003)
- **Ask more questions** (IDT interactivity – Hancock et al, 2007)
- **More 'I' words** (Ho et al, 2015 – anonymous)
- **Markers of cognitive complexity** (Hancock et al, 2008; Newman et al, 2003)
  - **Exclusive words** – *but, except, without* (less)
  - **Motion words** - *walk, move, go* (more)

Forensic linguistics

- **Authorship identification**
- Authorship characterisation
- **Similarity detection**

  - **Lexical:** # of words, words per sentence, word length, usage frequency
  - **Syntax:** punctuation, function words
  - **Structural:** greetings, signatures, paragraph length, # of paragraphs
  - **Content-specific:** key words
    - Abbasi & Chen, 2006

Case study VI: Diederik Stapel

- Analysed by Markowitz & Hancock, 2014
- More:
  - Scientific methods, certainty, emotional actions, states, processes
- Fewer adjectives, less descriptive
- Not a solution for identifying fraudulent research!

Passive detection

- Analysis – linguistic markers, some reliable indicators
- Undeutsch hypothesis (Undeutsch, 1967)
- Non-verbal behaviours (Tsikerdekis & Zeadally, 2014)

Active detection

- Increase cognitive load (move to synchronous, richer environments)
  - Request more sensory information, ask more questions
  - Emphasise elements underemphasised by deceiver
  - Introduce additional tasks (Vrij et al, 2008)
- Gaze detection as reliable indicator (Sporer & Schwandt, 2007; Zuckerman & Driver, 1985; Zuckerman et al, 1981)
  - Extended to video conferencing (Pak & Zhou, 2013)

Chapter IV: Fighting back

"It is double pleasure to deceive the deceiver."

Jean de La Fontaine, *Fables, II.15*

"But no pleasure is comparable to standing upon the vantage ground of truth."

Francis Bacon, *Of Truth*

Turning the tables

- Drip-feed false information (Heckman et al, 2015)
- Elicit information for use in attribution
- Assuming:
  - **Legal and ethical tests met**
  - **No conflict of interest with LEAs and other agencies**

Case study VII: Shannen Rossmiller

- Posed as an Al Qaeda affiliate
  - Older personae vouched for younger ones
  - Distinct personalities and backgrounds
  - ‘Martyred’ once no longer required
- Worked due to decentralised nature of AQ sympathisers

http://www.washingtonpost.com/wp-dyn/content/article/2006/06/03/AR2006060300530.html??noredirect=on

https://www.wired.com/2007/10/ff-rossmiller/

Chapter V: Conclusions

"One may outwit another, but not all the others."

François de La Rochefoucauld, *Maxim*, 394

ROSE detection checklist

- ❑ Have you met the person in real life?
- ❑ Has anyone you know met the person in real life?
- ❑ Reverse image search
- ❑ Does their knowledge check out?
- ❑ Independent verification of backgrounds and qualifications
- ❑ Has the person asked to contact you on your corporate email?
- ❑ Ask what the person wants, and why? Why from you?
- ❑ Are they interested in technical aspects?
- ❑ Are they evasive when asked to meet in person?
- ❑ What about a phone call or video call?

ROSE detection checklist

- ❑ Check for linguistic deception markers
- ❑ Similarity to other profiles (behavioural/linguistic/non-verbal)
- ❑ Response to increases in cognitive load
- ❑ Is conditioning behaviour being used?
- ❑ Marked interest in your job, industry, or research?
- ❑ Age of the profile. What's the earliest trace? *Why?*
- ❑ Inconsistencies in background, activity, or reactions
- ❑ Is there a lot of publicly available information on you?
- ❑ Have they asked you to whitelist anything?
- ❑ Do you have a way to report suspicious behaviour?

Future research

- Linking false personae
  - Methodology, linguistics, granular behaviours
  - Contact me if you want to get involved!
- Further research on online deception
- Detection of deepfakes and audio fakes
- Concept of personal filters and vulnerabilities
- Advances in perceptual hashing

Shameless plug

- If you're interested in human side-channels:
- **Come and see my DEF CON talk!**
- **Sunday @ 2pm, 101 Track**
- ***Betrayed by the keyboard: How what you type can give you away***

Black Hat Sound Bytes

1. ROSE is an insidious technique which can be very effective
2. Methods for detection are generally untested, but offer some hope
3. Detecting ROSE has benefits for society as a whole
4. Refer to the ROSE checklist, and let me know if you can expand it!

Feedback, questions, collaboration:

@darkartlab

matt.wixey@pwc.com

The end

"Fiction is like a spider's web, attached ever so lightly perhaps, but still attached to life at all four quarters. Often the attachment is scarcely perceptible."

Virginia Woolf, *A Room of One's Own*.

References

Abbasi, A., & Chen, H. (2006). Visualizing authorship for identification. In *International Conference on Intelligence and Security Informatics*.

Bachrach, J. (2005, February). "U want me 2 kill him?" *Vanity Fair*. Retrieved from: https://www.vanityfair.com/news/2005/02/bachrach200502

Barber, B. (1983). *The logic and limits of trust*. New Jersey: Rutgers University Press.

Berger, C. R., & Calabrese, R. J. (1975). Some exploration in initial interaction and beyond: Toward a developmental theory of communication. *Human Communication Research*, 1, 99–112.

Bond, C.F.. & DePaulo, B.M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review,* 10, 214-234

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6, 203–242.

Burgoon, J. K., Blair, J. P., Qin, T., & Nunamaker, J. F. (2003, June). Detecting deception through linguistic analysis. In *International Conference on Intelligence and Security Informatics* (pp. 91-101). Springer, Berlin, Heidelberg.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W. A., & Feldman, C. M. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communications Monographs*, *63*(1), 50-69.

Buss, D. M., Gomes, M., Higgins, D. S., & Lauterbach, K. (1987). Tactics of manipulation. *Journal of personality and social psychology*, *52*(6)

Carlson, J. R., George, J. F., Burgoon, J. K., Adkins, M., & White, C. H. (2004). Deception in computer-mediated communication. *Group Decision and Negotiation*, 13: 5-28.

Counter Threat Unit Research Team (2017). The curious case of Mia Ash: Fake persona lures Middle Eastern targets. *Dell Secureworks*. Retrieved from: https://www.secureworks.com/research/the-curious-case-of-mia-ash

Daft, R., & Lengel, R. (1986). Organizational information requirements, media richness, and structural design. *Management Science*, 32(5).

References

Daft, R., Lengel, R., & Trevino, L. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly*, 11(3).

Dasgupta, P. (2000). Trust as a commodity. In D. Gambetta (Ed.) *Trust: Making and breaking cooperative relations*: 49-72. New York: Blackwell.
Dennis, A., Fuller, R., & Valacich, J. (2008). Media, tasks, and communication processes: A theory of media synchronicity. *MIS Quarterly*, 32(3).

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 51–70). Netherlands: Kluwer.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979–995.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*(1), 74.

Drouin, M., Miller, D., Wehle, S. M., & Hernandez, E. (2016). Why do people lie online? "Because everyone lies on the internet". *Computers in Human Behavior*, *64*, 134-142.

Elsbach, K. D. (2004). Managing images of trustworthiness in organization. In R. M. Kramer & K. S. Cook (Eds.), *Trust and distrust in organizations: Dilemmas and approaches*. New York: Russell Sage Foundation.

Escritt, T., & Martina, M. (2017, December 10). German intelligence unmasks alleged covert Chinese social media profiles. *Reuters*. Retrieved from: https://www.reuters.com/article/us-germany-security-china/german-intelligence-unmasks-alleged-covert-chinese-social-media-profiles-idUSKBN1E40CA

FireEye. (2015). Behind the Syrian conflict's digital front lines. *FireEye*. Retrieved from: https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/rpt-behind-the-syria-conflict.pdf

Fogg, B. J. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. In *CHI'03 extended abstracts on human factors in computing systems*

George, J. F., Giordano, G., & Tilley, P. A. (2016). Website credibility and deceiver credibility: Expanding prominence-interpretation theory. *Computers in Human Behavior*, 54. http://dx.doi.org/10.1016/j.chb.2015.07.065

References

George, J. F., Marett, K., & Tilley, P. (2004). Deception detection under varying electronic media and warning conditions. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*.

Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Milton Keynes: Open University Press.

Good, D. (1988). Individuals, interpersonal relations, and trust. In D. Gambetta (Ed.) *Trust: Making and breaking cooperative relations:* 31-48. New York: Blackwell.

Granhag, P. A., Andersson, L. O., Stromwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: 'Criminal' beliefs about deception. *Legal and Criminological Psychology*, 9, 103–119.

Hancock, J. (2007). Digital Deception: When, where, and how people lie online. In K. McKenna, T. Postmes, U. Reips, & A. Joinson (Eds.), *Oxford handbook of internet psychology:* 287-301. Oxford: Oxford University Press.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*(1), 1-23.

Hancock, J. T., Curry, L., Goorha, S. & Woodworth, M. (2005). Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. *Proceedings of the 38th Hawaii International Conference on System Sciences*.

Harden, B. (2006, June 4). In Montana, casting a web for terrorists. *The Washington Post*. Retrieved from: http://www.washingtonpost.com/wp-dyn/content/article/2006/06/03/AR2006060300530.html??noredirect=on

Hartwig, M., Granhag, P. A., Stromwall, L. A., & Andersson, L. O. (2004). Suspicious minds: Criminals' ability to detect deception. *Psychology, Crime and Law*, 10, 83–94.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, *19*(4), 307-342.

Heckman, K. E., Stech, F. J., Schmoker, B. S., & Thomas, R. K. (2015). Denial and deception in cyber defense. *Computer*, *48*(4), 36-44.

Hitt, J. (2007, October 23). Behind enemy lines with a suburban counterterrorist. *Wired*. Retrieved from: https://www.wired.com/2007/10/ff-rossmiller/

References

Ho, S. M., Hancock, J. T., Booth, C., Liu, X., Timmarajus, S. S., & Burmester, M. (2015). Liar, Liar, IM on Fire: Deceptive language-action cues in spontaneous online communication, *IEEE Intelligence and Security Informatics:* 157-159.

Hu, X. S., Hong, K. S., & Ge, S. S. (2012). fNIRS-based online deception decoding. *Journal of Neural Engineering*, *9*(2), 026012.

Jarecki, A., Smerling, M., Joost, H., & Schulman, A. (Producers), & Joost, H. & Schulman, A. (Directors). (2010). *Catfish* [Motion picture]. United States: Universal Pictures.

Johnson, M. K. and C. L. Raye. (1981). Reality Monitoring. *Psychological Review* 88, 67–85.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237-251.

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law and Human Behavior*, *29*(2), 211-227.

Keila, P. S., & Skillicorn, D. B. (2005). Detecting unusual and deceptive communication in email. *External technical report*. School of Computing, Queen's University, Kingston, Ontario, Canada.

Knapp, M. L., & Comadena, M. A. (1979). Telling it like it isn't: A review of theory and research on deceptive communications. *Human Communication Research,* 5, 270–285.

Kottemann, K. L. (2015). *The rhetoric of deliberate deception: What catfishing can teach us*. PhD thesis, University of Louisiana.

Kramer, R. M. (1994). The sinister attribution error: Paranoid cognition and collective distrust in organizations. *Motivation and emotion*, *18*(2), 199-230.

Kramer, R. M., Brewer, M. B., & Hanna, B. A. (1996). Collective trust and collective action. *Trust in organizations: Frontiers of theory and research*, 357-389.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs*, 66, 125–144.

References

Magdy, W., Elkhatib, Y., Tyson, G., Joglekar, S., & Sastry, N. (2017). Fake it till you make it: Fishing for Catfishes. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*(pp. 497-504). ACM.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: police officers' ability to detect suspects' lies. *Journal of applied psychology*, *89*(1), 137.

Markowitz, D. M., & Hancock, J. T. (2014). Linguistic traces of a scientific fraud: The case of Diederik Stapel. *PloS one*, *9*(8).

Meissner, C. A., & Kassin, S. M. (2002). “He's guilty!”: Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*(5), 469-480.

Mischel, W. & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*, 102(2)

Mischel, W. (1999). Personality coherence and dispositions in a cognitive-affective personality system (CAPS) approach. In: D. Cervone & Y. Shoda (Eds.) *The coherence of personality: Social-cognitive bases of consistency, variability, and organization:* 37-60. New York: Guilford Press.

Newman, M. L., Pennebaker, J.W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29,* 665–675.

Niederhoffer, K. G., & Pennebaker, J.W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21,* 337–360.

Pak, J., & Zhou, L. (2013). Eye gazing behaviors in online deception. *AMCIS 2013 Proceedings*. Retrieved from http://aisel.aisnet.org/amcis2013/ISSecurity/RoundTablePresentations/3.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–225.

Pistone, J. (1988). *My Undercover Life in the Mob*. New York: Dutton.

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and criminological Psychology*, *15*(1), 57-75.

References

Powell, W. W. (1996). Trust-based forms of governance. In R.M. Kramer & T.R. Tyler (Eds). *Trust in organizations: Frontiers of theory and research*. California: Sage.

Ryan, T. (2010). Getting into bed with Robin Sage. *Provide Security*. Retrieved from: https://www.privacywonk.net/download/BlackHat-USA-2010-Ryan-Getting-In-Bed-With-Robin-Sage-v1.0.pdf

Shoda, Y., Mischel, W. & Wright, J.C. (1994). Intraindividual stability in the organization and patterning of behavior: incorporating psychological situations into the idiographic analysis of personality. *Journal of personality and social psychology*, 67(4)

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization science*, *4*(3), 367-392.

Sporer, S. L., & Schwandt, B. (2007) Moderators of nonverbal indicators of deception: A metaanalytic synthesis, *Psychology, Public Policy, and Law*, 13, 1, 1–34

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, *103*(2), 193-210

Toma, C. L., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, *62*(1), 78-97.

Tsikerdekis, M., & Zeadally, S. (2014). Multiple account identity deception detection in social media using nonverbal behavior. *IEEE Transactions on Information Forensics and Security*, *9*(8), 1311-1321.

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), *Handbuch der PsychologieVol. 11: Forensische Psychologie*: 26-181. Gottingen: Hogrefe.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice.* Chichester, England: Wiley

Vrij, A. (2008a). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: Wiley.

Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, B., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, 32, 253–265.

References

Zayas, V., Shoda, Y. & Ayduk, O.N. (2002). Personality in context: An interpersonal systems perspective. *Journal of personality*, 70(6), 851-900.

Zhou, L. (2005). An empirical investigation of deception behavior in instant messaging. *IEEE Transactions on Professional Communication*, *48*(2), 147-160.

Zhou, L., & Sung, Y. (2011). Discourse cues to online deception. In *Proceedings of the Credibility Assessment and Information Quality in Government and Business Symposium*.

Zhou, L., & Sung, Y. W. (2008). Cues to deception in online Chinese groups. In *Proceedings of the 41st Hawaii International Conference on System Sciences*.

Zhou, L., & Zhang, D. (2004). Can online behavior unveil deceivers? An exploratory investigation of deception in instant messaging. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences.*

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group decision and negotiation*, *13*(1), 81-106.

Zhou, L., Burgoon, J. K., Zhang, D., & Nunamaker, J. F. (2004). Language dominance in interpersonal deception in computer-mediated communication. *Computers in Human Behavior*, *20*(3), 381-402.

Zuckerman, M., & Driver, R. (1985) Effects of segmentation patterns on the perception of deception, *Journal of Nonverbal Behavior*, 9, 3, 160–168

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981) Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances In Experimental Social Psychology*: 1–59, New York: Academic Press