Triggering Empathy out of Malicious Intent: The Role of Empathy in Social Engineering Attacks

Verena Distler
verena.distler@unibw.de
University of the Bundeswehr
Munich, Germany

Yasmeen Abdrabou
yasmeen.essam@unibw.de
University of the Bundeswehr
Munich, Germany

Felix Dietz
felix.dietz@unibw.de
University of the Bundeswehr
Munich, Germany

Florian Alt
florian.alt@unibw.de
University of the Bundeswehr
Munich, Germany

ABSTRACT

Social engineering is a popular attack vector among cyber criminals. During such attacks, impostors often attempt to trigger empathy to manipulate victims into taking dangerous actions, for example, sharing their credentials or clicking on malicious email attachments. The objective of this position paper is to initiate a conversation on the tension between positive and negative aspects of empathy in HCI as it pertains to security-relevant behaviors. To this end, we focus on the malicious ways in which empathy can be instrumentalized in social engineering. We describe examples of such empathy-related social engineering attacks, explore potential solutions (including the automated detection of empathy-triggering communication, or of empathetic communication on the part of a potential victim), and discuss technical, social as well as organizational interventions. We highlight research challenges and directions for future work.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**.

KEYWORDS

Empathy, Social Engineering, Security

ACM Reference Format:
Verena Distler, Yasmeen Abdrabou, Felix Dietz, and Florian Alt. 2023. Triggering Empathy out of Malicious Intent: The Role of Empathy in Social Engineering Attacks. In *EmpathiCH workshop (EMPATHICH '23), April 23, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3588967.3588969

EMPATHICH '23, April 23, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0749-0/23/04.
https://doi.org/10.1145/3588967.3588969

1 EMPATHY — A DOUBLE-EDGED SWORD IN USABLE PRIVACY AND SECURITY?

Individuals, organizations, and societies increasingly depend on digital products and services to function. Security attacks on these products and services can have harmful consequences, including financial harm, personal harm (e.g., reputation), societal harm, and physical harm. Social engineering attacks are particularly effective attacks and exploit human psychology to trick the user into performing security-critical actions, for instance by providing confidential information to an attacker [40]. Social engineering attacks often attempt to manipulate the victim by, for example, triggering different emotions or by creating trust [4], persuading victims to take an action [27]. The attacks can happen through real-time, person-to-person contact (e.g., malicious phone calls), or they can happen without real-time communication with the attacker (e.g., via email, so-called phishing). Social engineering attacks are a particularly relevant threat, often leading to harmful follow-up attacks, such as ransomware. Ransomware is software that denies access to a system or information until a ransom is paid [4], and often has highly disruptive and harmful consequences [29, 45].

Social engineering attacks frequently use persuasive design, including authority, social proof, liking, and reciprocity [18]. While the traditional persuasive principles do not specifically mention triggering empathy, many successful social engineering attacks ask victims for help in an attempt to trigger empathy and trigger a behavioral reaction (see section 2). While empathy is usually associated with pro-social behavior, it can also be exploited to manipulate people [9]. For instance, while cognitive empathy could be used to understand others better, and help them, cognitive empathy could also be used to gain an advantage over competitors in a business context, or intentionally embarrass another person [21]. In the case of social engineering, attackers can try to intentionally trigger empathy in a victim to encourage them to take action that can help the attacker.

The concept of empathy, and its role in HCI, is subject to much debate. There are a variety of definitions of empathy, which commonly distinguish between cognitive and affective empathy [13]. Cognitive empathy is concerned with people's ability to *understand* others' mental states (e.g., beliefs, intentions, emotions) [30]. In comparison, affective empathy involves *affect* on the part of the person who empathizes [30], focusing more on the experience of

emotion [9]. There is a debate about whether empathy is a relatively stable, trait-like concept, or a situation-dependent state [9]. The association between empathy and behavior is also subject to debate. Evidence suggests that empathy can lead to behavioral responses, but this is not always the case, and behavior can be mediated through other factors [9]. Empathy can thus be conceptualized as a motivational factor for behavior [9]. Empathy is considered a key factor in user-centered design, where "empathic design" intends to help designers empathize with users [13]. The idea that designers can easily "empathize" even with marginalized communities that they are not a part of themselves has been heavily criticized [37].

This position paper explores the idea that, if we were able to recognize that communication from a supposed attacker to the victim intends to trigger empathy, we could intervene. Similarly, if we were able to detect that a victim has an empathetic reaction to a message that we cannot clearly authenticate, we might be able to intervene. For example, by using friction design, we might be able to give a potential victim the space to reflect on an empathy-inducing request from an attacker, and consciously consider taking action [12]. Indeed, to a certain extent, people are able to regulate their empathetic response to stimuli, but it is often challenging to hit the ideal level of empathy [21]. For instance, "over-shooting" in terms of emotional empathy can lead to personal distress, "overshooting" in terms of cognitive empathy can lead to a loss of self. Having a relatively low level of emotional empathy can lead to a lack of sensitivity, a low level of cognitive empathy can lead to ego-centrism [21].

Summary. Empathy can have a positive influence on security-related behaviors, for instance as empathy is often a motivating factor for pro-social behavior [9] (e.g., helping a friend adopt more secure behaviors online out of empathy). However, malicious actors can also trigger empathy for purposes such as social engineering attacks. This tension calls for a nuanced approach to empathy in usable privacy and security (UPS) to help potential victims avoid falling for manipulative communications on the part of the attacker.

This paper brings a UPS perspective to the topic of empathy. The objective of this position paper is to initiate a conversation on the tension between positive and negative aspects of empathy in HCI as it pertains to security-relevant behaviors. We exemplify this tension with scenarios in which empathy is used for social engineering (section 2). We then explore potential solutions to detect and counteract the intentional triggering of empathy for malicious purposes on a technical, social and policy level (section 3), and highlight research challenges (section 4).

2 SAMPLE SCENARIOS WHERE EMPATHY IS USED FOR SOCIAL ENGINEERING

The following scenarios serve to exemplify instances where empathy is triggered out of malicious intent. In many of these scenarios, malicious actors exploit a person's willingness to help another person in need. These scenarios are not meant to provide an exhaustive list of empathy-related social engineering attacks. They were selected to illustrate the way empathy-related social engineering attacks can take place on a variety of platforms and technologies.

Instrumentalizing empathy in telephone scams. Empathy is often triggered intentionally in telephone scams. Some of these scams target elderly people by posing as a relative or friend who is in desperate need of money for example because they pretend to be involved into an accident. By appealing to their emotions and instilling a sense of urgency, the scammer convinces victims to hand over cash or valuable items [1]. Another type of telephone scam can be used to gain information in organizational contexts, where a caller might impersonate another employee of the company and try to persuade the victim (e.g., administrative staff) to disclose sensitive information.

Instrumentalizing empathy in phishing attacks. Utilizing the internet as a medium, phishing or scam emails can trick victims into entering their credentials of (bank) accounts into forged websites or even directly wiring money to fraudsters. Often, these attacks use a forged personal story to create empathy and promise to pay everything back when their issues are resolved [2]. These scams can potentially be exposed afterwards by talking to the supposed relative or friend or asking the bank if the payment recipient was legitimate.

Instrumentalizing empathy in donation scams. Requests for donations, for instance in the case of natural disasters, often attempt to trigger empathy. It can be difficult to verify who is really asking for donations, and scams are frequent. In these cases, victims could receive multiple donation requests for the most current natural disaster without ever realizing that the money does not reach the intended destination [3].

Instrumentalizing empathy in the phyiscal world. Empathy-related social engineering attacks do not only take place in digital spaces. In the physical world, a person could provide assistance to someone in need, while a third person takes advantage of their distraction and steals something from person helping. Here, thieves can purposefully create situations that trigger empathy to preoccupy good Samaritans and turn them into victims.

3 EXPLORING SOLUTIONS

In the previous sections, we highlighted how empathy could be used to manipulate users in the context of social engineering attacks. In this section, we reflect on some possible solutions for countering the use of empathy to perform a social engineering attack.

These solutions differ on multiple dimensions. Raising awareness (section 3.1) through training interventions typically intends to create knowledge about potential attacks *before* a potential victim is exposed to such an attack. Training interventions do not require the ability to detect a threat in the moment when it occurs, but have their limitations. Improved knowledge does not always lead to behavioral outcomes when a threat occurs. Thus, interventions that are triggered when a threat occurs are promising, for instance warnings. Such in-the-moment interventions require the ability to technically detect a threat to avoid a large number of unnecessary warnings. We explore options to detect threats in section 3.2. In

[1] https://www.cprcallblocker.com/blogs/news/grandparents-in-uk-warned-to-hang-up-on-new-phone-scam
[2] https://www.aura.com/learn/wire-transfer-scams
[3] https://consumer.ftc.gov/features/how-donate-wisely-and-avoid-charity-scams

section 3.3, we describe possible interventions, on a technical, social and policy level.

3.1 Raise Awareness

Raising users' awareness about the different techniques that could be used to manipulate them is the first step to reducing their vulnerability to social engineering attacks. This can take the form of training that includes putting users in empathetic scenarios, or role-playing and teaching them how to spot such attacks. It also seems promising for users not only to play the victim, but also the attacker, to better reflect on how they can be targeted and which traits others might target to put them at risk. Raising awareness can also aim at increasing users' self-efficacy in different ways by, for example, using serious games [2]. Training interventions must be carefully designed to avoid unintended outcomes, such as learned helplessness, a phenomenon where individuals believe situations are out of their control. [41].

3.2 Detecting Empathy-Related Threats

Although raising awareness is an important first step, literature has shown that training can not work on its own as training has a short-term effect and has to be repeated frequently [5]. In addition, improved knowledge does not necessarily lead to behavior change. Hence, training has to be accompanied by a way to communicate to users at the right time that somebody might be trying to manipulate them. This moment might be when a potential victim is reading on social media, receiving an email, or even a phone call, each of which might be a threat intending to maliciously trigger an empathetic response. To enable threat communication, we must first detect the threat and find suitable ways to communicate it. Below we point out different directions that have shown to be promising in various fields. It seems promising to investigate these techniques further.

As empathy includes affective (emotions) and cognitive (attitudes) components [6] we can investigate the usage of metrics from both domains. Such metrics can include voice, text, facial expressions, and physiological signals.

Voice. Voice has long been used for emotion detection. If we consider empathy an affect state, then voice might be used to detect empathy-related affect on the part of a potential victim. Existing work shows the ability to detect empathy from users' voices. For example, Alam et al. [3] conducted a study to detect empathy in spoken conversations in call centers. The authors considered acoustic, lexical, and psycholinguistic features for their generated model. Other works use paralinguistic vocal cues such as pitch, cadence, speed, and volume [25]. Overall, voice linguistics shows great promise for conveying and perceiving empathy. Additional work is needed in the context of social engineering linguistics both for detecting an attack attempt from the attacker and to warn users if empathy was detected on their end. We believe this could be helpful to protect users in the context of voice calls (i.e. vishing) and videos on social media (i.e. fake content intending to trigger empathy).

Text. Written text is another possibility to convey and trigger empathy. Text-based detection of empathy, or empathy-triggering conversations, can be applied to both sides; to detect (1) whether the attacker is triggering empathy and (2) whether the potential victim's empathy has been triggered. However, to detect empathy from text, it is necessary to have a database containing empathetic exchanges. Several databases exist online based on empathetic conversations [28, 35, 36, 46], making detection easier. Moreover, Quintanilla et al. [17] showed participants a set of stories to study the potential of linguistic choices and patterns in the stories to trigger empathy with the characters in the narrative. The findings describe the interaction between narrative technique and readers' evaulation of the moral of the characters. The authors found that references to dictatorship or real-world references seemed to trigger to empathy. We believe that using existing datasets in the literature along with detecting repeated patterns can be a promising direction to detect empathy in conversational text [22, 32]. Additional training on the linguistics of social engineering empathy could help users avoid falling for e.g. online scams, phishing emails, and fake news.

Facial Expressions. Facial expressions might reflect people being exposed to empathy-inducing situations [14, 16]. For example, pulling down eyebrows in a flat way and pointing them forward over the bridge of the nose is a sign of sympathy. Similarly, not pulling eyelids tight or raised, head and body-oriented forward, bottom eyelids raised slightly, and lower face relaxed are also signs of sympathy [16]. Beyond sympathy detection, in literature, facial expressions were found to correlate with users' emotions [24], task performance [42], cognitive profiles [15], and cognitive load [23]. Although the interpretation of facial expressions is controversial [7] and sometimes only reflect negative or positive emotions as a whole without subdivisions, it might be interesting to investigate the potential of facial expressions further, especially if combined with other measures such as voice or text.

Physiological Signals. There have been attempts to use physiological signals to detect empathy. For example, Kumano et al. [26] researched emotional interactions in meetings. The authors targeted empathy as one aspect and used eye gaze, facial expressions, and speech-silence features for detection. Using Markov models, the authors found that gaze direction, mutual gaze, and facial expression patterns to be promising for detecting empathy. It is also worth mentioning that gaze movements as a physiological aspect are being used heavily as metrics to protect users from social engineering attacks [1, 10]. However, gaze tracking is just one physiological aspect; electroencephalography (EEG) signals have also been used to detect empathy in VR [38]. Salminen et al. developed a social biofeedback VR environment for the conducting of simplified empathy exercises, that are inspired by traditional meditation practices. The authors found statistically significant positive correlations between perceived empathy and frontal asymmetry and respiratory rate. Similarly, skin temperature has shown close relation to empathy in certain scenarios. For example, Moline et al. [31] found large temperature changes in highly-empathetic participants. On the other side, the low-empathy participants skin temperature change was almost always non-significant.

Of course, none of these studies provide the full answer to how to detect empathy-inducing communication, or how to detect empathy in a potential victim. However, we see these studies as interesting first steps that show the potential of exploring the detection of malicious uses of empathy in social engineering attacks.

3.3 Interventions

3.3.1 Technical Interventions. After highlighting the different ways that could be used to detect empathy and empathy-inducing communication, the next step is to nudge or alert users from possible attacks. Nudging has shown to be a promising technique for adjusting users' behavior. It can also take different forms, such as visual, haptic, and auditive nudges. Different research exists on nudging users not to fall for phishing emails, for example, adding a warning near the phishing link [33] and many more discussed by Caraban et al. [8]. However, one should consider the trade-off between cost and benefit regarding user nudging [19]. Users tend to get nudge fatigue, and hence the nudge design should change over time, whether in size, color, or position to eliminate fatigue effect [20, 34, 43, 44]. It is also important to accommodate for false positives and false negatives. Such false positive or false negative nudges can put users at risk or annoy them, thus negatively affecting the overall user experience.

3.3.2 Social Interventions. Going beyond technical interventions such as nudges and warnings, it is also promising to encourage a social approach to help potential victims be less vulnerable to empathy-related social engineering attacks. Beyond raising awareness by informing at-risk populations about social engineering attacks, municipalities could encourage them to call a help line in cases where they are unsure of how to react to a potential attack. The helpline could provide strategies to authenticate a caller or potential attacker, and help the potential victim delay action even in response to seemingly urgent requests from the attacker. Forums or group chats could also be used to get timely advice from others about how to react to a potential social engineering attack.

3.3.3 Organizational policies. Beyond raising general awareness, organizations could help employees resist to empathy-related social engineering attacks by providing, for instance, flow-charts on how to authenticate a caller asking for sensitive information, and how to delay action in uncertain cases. Organizations could also introduce policies on the type of information that can be provided by phone or email, and which communication channels should be used for sensitive information (e.g., secure messaging).

4 CHALLENGES

We have argued that, in the context of social engineering, it is worthwhile to try and detect and counteract attempts to maliciously trigger empathy. Various research challenges remain.

Classification Accuracy and Measures of Empathy. To counteract potential victims taking harmful actions, it seems promising to warn them in cases when we can reasonably assume that an attack is underway. To do so, we would need to be able to detect, for instance, empathy in the potential victim, as well as potentially malicious communication that triggers empathy. We have described potential solutions in section 3.2, but future work needs to investigate these further, and empirically study how well suited these methods are to infer empathy. It seems especially promising to investigate ways of detecting communication that intends to trigger empathy, for instance in communications that do not stem from a trusted source. A related open challenge is how to deal with false positives, leading to potential victims being overloaded with warnings which they might start to ignore [39].

A related challenge is how to measure empathy, both on a measurement level (e.g., in voice, text, physiology) and as a self-reported measure. As we previously mentioned, there are different attempts to quantify and calculate measured empathy scores, and context plays an important role. Moreover, until now, we lack a generally accepted self-reported measurement of empathy to refer to, and further work is needed to continue improving empathy measures including self reflections, questionnaires, and studying physiological responses.

Ethical Considerations. Social engineering studies involving people often depend on deception to conceal the main aim of the study. The resulting tension between realistic risk representation and ethical, practical and legal concerns is typical for studies in the field of UPS [11]. Researchers should carefully weigh ethical trade-offs when conducting studies with people that involve the manipulation of empathy.

Finding the Sweet Spot. In most communications that trigger empathy, it is justified to feel empathetic, and perhaps take the action of helping the person (e.g., a colleague asking for help with a security-related question). In other cases, an attacker might try to trigger empathy in a potential victim, and there is reasonable cause for doubt (an unknown sender of an email, a new number calling, an urgent or authoritative communication style). In these cases, we should encourage potential victims to carefully consider whether they should take the action that is asked of them. An open research challenge is how to strike the right balance between encouraging empathy, and introducing friction to encourage deliberate reflection [12].

5 CONCLUSION

Empathy is used in many successful social engineering attacks. While empathy is only one of the motivational factors that contribute to a behavioral response, it is worthwhile to investigate further (1) how empathy is used in social engineering to manipulate potential victims (2) how communication intending to trigger empathy might be detected, and how empathetic response might be detected (3) which interventions might be helpful to help potential victims avoid falling for empathy-related social engineering attacks. By bringing a security perspective to the topic of empathy, this contribution creates an opportunity for a conversation on malicious uses of empathy, and how we might help potential victims avoid manipulation.

ACKNOWLEDGMENTS

The presented work received funding from the German Research Foundation (DFG) under project no. 316457582 and from the Studienstiftung des Deutschen Volkes. This research was also funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr (Voice of Wisdom). dtec.bw is funded by the European Union – NextGenerationEU.

REFERENCES

[1] Yasmeen Abdrabou, Elisaveta Karypidou, Florian Alt, and Mariam Hassib. 2023. Investigating User Behaviour Towards Fake News on Social Media Using Gaze

and Mouse Movements. In *Proceedings of the Usable Security Mini Conference 2023 (USEC'23)*. Internet Society, San Diego, CA, USA. https://doi.org/10.14722/usec.2023.232041

[2] Dina Aladawy, Kristian Beckers, and Sebastian Pape. 2018. PERSUADED: Fighting Social Engineering Attacks with a Serious Game. In *Trust, Privacy and Security in Digital Business*, Steven Furnell, Haralambos Mouratidis, and Günther Pernul (Eds.). Springer International Publishing, Cham, 103–118.

[3] Firoj Alam, Morena Danieli, and Giuseppe Riccardi. 2016. Can we detect speakers' empathy?: A real-life case study. In *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 000059–000064.

[4] Hussain Aldawood and Geoffrey Skinner. 2020. An Advanced Taxonomy for Social Engineering Attacks. *International Journal of Computer Applications* 177, 30 (Jan. 2020), 1–11. https://doi.org/10.5120/ijca2020919744

[5] Abdullah Alnajim and Malcolm Munro. 2009. An Evaluation of Users' Anti-Phishing Knowledge Retention. In *2009 International Conference on Information Management and Engineering*. 210–214. https://doi.org/10.1109/ICIME.2009.114

[6] Simon Baron-Cohen and Sally Wheelwright. 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders* 34 (2004), 163–175.

[7] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68. https://doi.org/10.1177/1529100619832930 arXiv:https://doi.org/10.1177/1529100619832930 PMID: 31313636.

[8] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300733

[9] Benjamin M.P. Cuff, Sarah J. Brown, Laura Taylor, and Douglas J. Howat. 2016. Empathy: A Review of the Concept. *Emotion Review* 8, 2 (April 2016), 144–153. https://doi.org/10.1177/1754073914558466

[10] A. Darwish and E. Bataineh. 2012. Eye tracking analysis of browser security indicators. In *2012 International Conference on Computer Systems and Industrial Informatics*. 1–6.

[11] Verena Distler, Matthias Fassl, Hana Habib, Katharina Krombholz, Gabriele Lenzini, Carine Lallemand, Lorrie Faith Cranor, and Vincent Koenig. 2021. A Systematic Literature Review of Empirical Methods and Risk Representation in Usable Privacy and Security Research. *ACM Trans. Comput.-Hum. Interact.* 28, 6, Article 43 (dec 2021), 50 pages. https://doi.org/10.1145/3469845

[12] Verena Distler, Gabriele Lenzini, Carine Lallemand, and Vincent Koenig. 2020. The Framework of Security-Enhancing Friction: How UX Can Help Users Behave More Securely. In *New Security Paradigms Workshop 2020*. ACM, Online USA, 45–58. https://doi.org/10.1145/3442167.3442173

[13] Luce Drouet, Kerstin Bongard-Blanchy, Vincent Koenig, and Carine Lallemand. 2022. Empathy in Design Scale: Development and Initial Insights. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–7. https://doi.org/10.1145/3491101.3519848

[14] Nancy Eisenberg, Richard A Fabes, Paul A Miller, Jim Fultz, Rita Shell, Robin M Mathy, and Ray R Reno. 1989. Relation of sympathy and personal distress to prosocial behavior: a multimethod study. *Journal of personality and social psychology* 57, 1 (1989), 55.

[15] Mohamed El Kerdawy, Mohamed El Halaby, Afnan Hassan, Mohamed Maher, Hatem Fayed, Doaa Shawky, and Ashraf Badawi. 2020. The automatic detection of cognition using eeg and facial expressions. *Sensors* 20, 12 (2020), 3516.

[16] Caroline J Falconer, Janek S Lobmaier, Marina Christoforou, Sunjeev K Kamboj, John A King, Paul Gilbert, and Chris R Brewin. 2019. Compassionate faces: Evidence for distinctive facial expressions associated with specific prosocial motivations. *PloS one* 14, 1 (2019), e0210283.

[17] Carolina Fernandez-Quintanilla. 2020. Textual and reader factors in narrative empathy: An empirical reader response study using focus groups. *Language and Literature* 29, 2 (2020), 124–146. https://doi.org/10.1177/0963947020927134 arXiv:https://doi.org/10.1177/0963947020927134

[18] Ana Ferreira, Lynne Coventry, and Gabriele Lenzini. 2015. Principles of Persuasion in Social Engineering and Their Use in Phishing. In *Human Aspects of Information Security, Privacy, and Trust (Lecture Notes in Computer Science, Vol. 9190)*, Theo Tryfonas and Ioannis Askoxylakis (Eds.). Springer International Publishing, Cham, 36–47.

[19] Jeff French. 2011. Why nudging is not enough. *Journal of Social Marketing* (2011).

[20] Pelle Guldborg Hansen and Andreas Maaløe Jespersen. 2013. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation* 4, 1 (2013), 3–28.

[21] Sara D. Hodges and Robert Biswas-Diener. 2007. Balancing the empathy expense account: strategies for regulating empathic response. In *Empathy in Mental Illness* (1 ed.), Tom F. D. Farrow and Peter W. R. Woodruff (Eds.). Cambridge University Press, 389–407. https://doi.org/10.1017/CBO9780511543753.022

[22] Mahshid Hosseini and Cornelia Caragea. 2021. Distilling knowledge for empathy detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. 3713–3724.

[23] M Sazzad Hussain, Rafael A Calvo, and Fang Chen. 2014. Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference. *Interacting with computers* 26, 3 (2014), 256–268.

[24] Christian G. Kohler, Travis Turner, Neal M. Stolar, Warren B. Bilker, Colleen M. Brensinger, Raquel E. Gur, and Ruben C. Gur. 2004. Differences in facial expressions of four universal emotions. *Psychiatry Research* 128, 3 (2004), 235–244. https://doi.org/10.1016/j.psychres.2004.07.003

[25] Michael W Kraus. 2017. Voice-only communication enhances empathic accuracy. *American Psychologist* 72, 7 (2017), 644.

[26] Shiro Kumano, Kazuhiro Otsuka, Dan Mikami, and Junji Yamato. 2011. Analyzing empathetic interactions based on the probabilistic modeling of the co-occurrence patterns of facial expressions in group meetings. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 43–50.

[27] Stephen EG Lea, Peter Fischer, and Kath M Evans. 2009. The psychology of scams: Provoking and committing errors of judgement. (2009).

[28] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957* (2017).

[29] Sean Lyngaas. 2022. Ransomware attack hits New Jersey county. https://www.cnn.com/2022/05/26/politics/new-jersey-somerset-county-ransomware-attack/index.html

[30] Heidi L. Maibom. 2017. *The Routledge Handbook of Philosophy of Empathy* (1 ed.). Routledge, New York : Routledge, 2017. |. https://doi.org/10.4324/9781315282015

[31] A Moliné, J Fernández-Gómez, E Moya-Pérez, M Puertollano, G Gálvez-García, O Iborra, and E Gómez-Milán. 2018. Skin temperature reveals empathy in moral dilemmas: An experimental thermal infrared imaging study. *Thermology International* 28, 4 (2018), 197–206.

[32] Edwin Carlos Montiel-Vázquez, Jorge Adolfo Ramírez Uresti, and Octavio Loyola-González. 2022. An Explainable Artificial Intelligence Approach for Detecting Empathy in Textual Communication. *Applied Sciences* 12, 19 (2022). https://doi.org/10.3390/app12199407

[33] Justin Petelka, Yixin Zou, and Florian Schaub. 2019. Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300748

[34] Aditya Kumar Purohit and Adrian Holzer. 2019. Functional Digital Nudges: Identifying Optimal Timing for Effective Behavior Change. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312876

[35] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. I know the feeling: Learning to converse with empathy. (2018).

[36] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207* (2018).

[37] Rebecca Rouse. 2021. Against the Instrumentalization of Empathy: Immersive Technologies and Social Change. In *Augmented and mixed reality for communities* (first edition ed.), Joshua A. Fisher (Ed.). CRC Press, Taylor & Francis Group, Boca Raton.

[38] Mikko Salminen, Simo Järvelä, Antti Ruonala, Ville J. Harjunen, Juho Hamari, Giulio Jacucci, and Niklas Ravaja. 2022. Evoking Physiological Synchrony and Empathy Using Social VR With Biofeedback. *IEEE Transactions on Affective Computing* 13, 2 (2022), 746–755. https://doi.org/10.1109/TAFFC.2019.2958657

[39] Angela Sasse. 2015. Scaring and Bullying People into Security Won't Work. *IEEE Security & Privacy* 13, 3 (May 2015), 80–83. https://doi.org/10.1109/MSP.2015.65

[40] Chandra Sekhar Bhusal. 2021. Systematic Review on Social Engineering: Hacking by Manipulating Humans. *Journal of Information Security* 12, 01 (2021), 104–114. https://doi.org/10.4236/jis.2021.121005

[41] Martin EP Seligman. 1972. Learned helplessness. *Annual review of medicine* 23, 1 (1972), 407–412.

[42] Kshitij Sharma, Evangelos Niforatos, Michail Giannakos, and Vassilis Kostakos. 2020. Assessing Cognitive Performance Using Physiological and Facial Features: Generalizing across Contexts. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 95 (sep 2020), 41 pages. https://doi.org/10.1145/3411811

[43] Cass R. Sunstein. 2017. Nudges that fail. *Behavioural Public Policy* 1, 1 (2017), 4–25. https://doi.org/10.1017/bpp.2016.3

[44] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A Field Trial of Privacy Nudges for Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2367–2376. https://doi.org/10.1145/2556288.2557413

[45] Leah Zhang-Kennedy, Hala Assal, Jessica Rocheleau, Reham Mohamed, Khadija Baig, and Sonia Chiasson. 2018. The aftermath of a crypto-ransomware attack at a large academic institution. In *27th ${[$USENIX$]$ Security Symposium (${[$USENIX$]$ Security 18)*. 1061–1078.

[46] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.