Human Systems Engineering and Design (IHSED 2026), Vol. 237, 2026, 736–744
https://doi.org/10.54941/ahfe1008212

Cognitive Model of Spear Phishing Cyberattacks

Béatrice Foucault[1] and Raphaël Gontran[2]

[1]Orange Research, France
[2]Université Toulouse Jean Jaurès, France

ABSTRACT

This research proposes a systemic theoretical model to understand the effectiveness of spear phishing in the context of fake banking advisor fraud, which affects 20% of the French population. Based on in-depth cognitive interviews, the study reveals how attackers orchestrate a cascade of psychological manipulations in four interdependent phases, systematically exploiting universal cognitive biases and emotional responses to bypass victims' analytical reasoning. The results demonstrate that vulnerability to spear phishing is not simply due to individual negligence but results from sophisticated exploitation of fundamental cognitive mechanisms, leading to a gradual deterioration of judgment and decision-making abilities. This detailed understanding of the underlying processes suggests a paradigm shift towards preventive strategies that emphasize the development of "situational intelligence" and protective metacognition, repositioning humans not as the "weak link" but as aware actors capable of recognizing their own cognitive vulnerabilities and activating appropriate protective strategies against digital manipulations.

Keywords: Spear phishing, Social engineering, Banking fraud, Cognitive biases, Cognitive cybersecurity, Human factors, Systemic model

INTRODUCTION

Context and Issue

You may have had this unfortunate experience? Phone scams are multiplying. ARCEP, the telecom regulator in France, has noted a surge in reports for this reason on its platform. The number of complaints increased from 531 in 2023 to over 19 000 in 2025.

Among these threats, spear phishing stands out due to its high level of sophistication and formidable effectiveness. Unlike traditional phishing, which often involves mass dissemination of generic messages via email or SMS, spear phishing targets specific individuals, relying on meticulous collection of personal information and advanced social engineering techniques (Kaspersky, 2025). This personalization makes the attack particularly difficult to detect and thwart.

The scale of the phenomenon is confirmed by a recent study (OpinionWay for Orange, 2025) concerning fake banking advisor fraud. Indeed, 20% of French people report having been victims, with or without financial loss.

Received March 26, 2026 Revised June 15, 2026 Accepted June 28, 2026 Available online September 25, 2026

This figure illustrates the vulnerability of the public to techniques that subtly exploit our emotions, habits, and cognitive weaknesses.

The modus operandi of this fraud follows a well-rehearsed scenario: an individual impersonates a bank advisor or security officer and contacts their target by phone. Claiming an urgent situation, such as attempted fraud on the account, a suspicious transaction, or an imperative security update, they create a sense of temporal and emotional pressure. The fraudster then capitalizes on the institutional trust placed in the bank to extract confidential information (access codes, bank card numbers) or to persuade the victim to transfer funds to accounts controlled by criminals (Krombholz et al., 2014).

Research Objectives

Our study aims to address a fundamental question: what are the underlying cognitive mechanisms that make spear phishing so effective, and how can prevention be improved by leveraging this understanding?

To answer this, we have developed a systemic approach that simultaneously illuminates the sequence of malicious actions orchestrated by the attacker, the mental processes activated in the target at each stage, and the cognitive biases exploited throughout the attack.

Methodology

Building on a state of the art dedicated to cognitive cybersecurity, our research employs in-depth cognitive interviews (Demarchi & Py, 2006), which facilitate the recall of more precise and comprehensive memories. These interviews were conducted with seven victims of fake banking advisor fraud. This qualitative approach allows for an accurate reconstruction of the attack from the victim's perspective, identification of decisive turning points, and illumination of the decision-making processes and emotions involved.

The analysis of the interviews, carried out using the grounded theory method (Paillé, 2017), revealed the recurrence of a narrative structure within the accounts and led to the development of a theoretical model articulated into four distinct phases, each activating specific psychological levers. This model was subsequently validated by cybersecurity experts.

DESCRIPTION OF THE SYSTEMIC MODEL

Our model, called the "CMAA Model", conceptualizes the fake banking advisor attack via spear phishing as an empirical, sequential, and dynamic process, comprising four interdependent phases: Capture, Manipulation, Action, and Awareness (Figure 1). Each phase of the model corresponds to specific tactical objectives for the attacker and involves particular cognitive mechanisms in the target.

Before addressing the actual attack scheme, it is important to mention a preliminary phase, known as the "pre-attack." This stage corresponds to the active preparation period of the attacker and holds strategic significance because it enables the collection of targeted information, which is essential

for establishing a climate of trust between the attacker and the target. It often involves weeks of research using public data, social networks, and other sources to increase the likelihood of success.

On the side of the target and future victim, the emotional and situational context is equally crucial for understanding the situation. Whether immobilized in traffic, overwhelmed by work at the office, absorbed by a crucial football match, or busy at home with children, stress and cognitive overload can impair vigilance. Indeed, these factors, stress, cognitive overload, emotional context, contribute to inattentional blindness, reducing an individual's ability to detect peripheral signals when focused on a primary task (Livet, 2016). One victim even stated: "In other circumstances, it wouldn't have happened like this".

The attack typically occurs via a standard phone call, sometimes preceded by a false call purportedly from the gendarmerie or police, intended to legitimize the subsequent intervention of the so-called banking advisor.

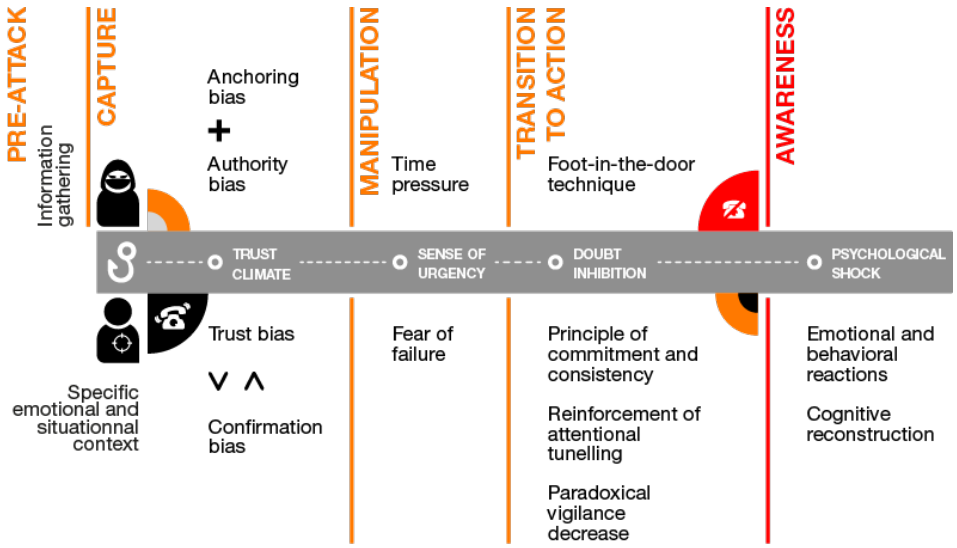

Figure 1: CMAA model.

Phase 1: Initial Contact and Attention Capture

The first phase of the attack involves establishing a climate of trust to legitimize the initial contact. The attacker presents themselves as recognized authority by impersonating a fraud/security officer from the victim's bank or from the Visa/MasterCard network. They often use the technique of "Caller ID Spoofing" (GSMA, 2023), which involves falsifying the official phone number of the bank. This deception is usually accompanied by the disclosure of specific personal information previously collected, such as the full name, part of the account number, and even details of recent transactions, giving the exchange an appearance of legitimacy and enhancing the attacker's credibility.

This phase activates several powerful cognitive biases:

Authority Bias: The foundational work of Milgram in 1963 (Guéguen, 2010) demonstrated our tendency to obey figures of authority, even when it conflicts with our moral judgment. In the context of fake banking advisor fraud, simply mentioning an official title ("fraud department manager") triggers this bias, reducing the victim's critical thinking (Cialdini, 2021).

Anchoring Bias: By providing accurate personal information early in the exchange, the attacker creates a "cognitive anchor" that guides the victim's subsequent reasoning toward perceiving the interlocutor as legitimate.

Confirmation Bias: Once initial trust is established, the victim tends to interpret subsequent information as confirming the caller's legitimacy, while ignoring or minimizing potential warning signals that may arise in their mind.

Truth Bias: Our natural tendency to accept statements as true by default (Gilbert et al., 1993) facilitates the acceptance of the attacker's discourse, especially when it relies on verifiable factual elements.

At this stage, the victim activates a heuristic processing mode rather than an analytical one. This thinking system is fast, intuitive, emotional, and automatic (System 1 of Kahneman, 2011). Faced with an unexpected call from a trusted institution, they rely on mental shortcuts ("My bank is calling to warn and protect me") rather than critically analysing the situation. The reassuring and professional tone of the attacker reinforces this cognitive orientation.

Phase 2: Cognitive and Emotional Manipulation

Once trust is established, the attacker shifts to creating a sense of urgency and an imminent threat (Longtchi et al., 2022). They typically announce the detection of ongoing fraudulent transactions on the victim's account, involving substantial amounts. The goal is to bypass rational reasoning by triggering an intense emotional response. The attacker encourages the victim to act quickly for fear that it will be too late.

The victim then experiences a feeling of fear. Fear is a primary emotion that triggers automatic survival responses (LeDoux, 1996). In the face of the threat of significant financial loss, the victim's limbic system activates, reducing activity in the prefrontal cortex responsible for analytical reasoning. Under this stress, the victim's attention becomes solely focused on the announced threat, at the expense of peripheral signals that might raise suspicion (Easterbrook, 1959). Thus, the narrowing of attentional focus prevents a comprehensive assessment of the situation.

Two types of levers are used by the attacker:

Loss Aversion: The work of Kahneman and Tversky (1979) on prospect theory demonstrated that the psychological pain of a loss is about twice as intense as the pleasure of an equivalent gain. Reasoning is thus short-circuited by emotions, leading to poor decision-making. The attacker exploits this bias by framing the situation as a potential loss to be avoided rather than a gain to be achieved (Chelsea et al., 2021).

Time Scarcity: By emphasizing the urgent and limited nature of the intervention ("we must act immediately"), the attacker activates the scarcity bias, which increases the perceived value of the proposed action and diminishes critical reflection (Cialdini, 2021).

This phase causes a significant degradation in decision quality. The victim shifts from a state of normal vigilance to acute stress, characterized by a reduction in available working memory and difficulty in considering alternative scenarios. Their reasoning becomes solely focused on immediate problem resolution. Paradoxically, their trust in the interlocutor, perceived as a "saviour", increases.

Phase 3: Action and Doubt Inhibition

This third phase aims to obtain from the victim the execution of a series of concrete actions: sharing security codes, validating transactions, transferring funds to a "secure account", etc. The attacker orchestrates a gradual escalation of commitments, where each action facilitates the next.

Several techniques are employed here to divert the victim's attention:

Principle of Commitment and Consistency: Once a person has made an initial decision or taken a first action, they tend to maintain consistency with that initial choice, even in the face of contradictory information (Festinger, 1957). Each completed action reinforces the victim's engagement in the process.

Foot-in-the-Door Technique: The attacker begins with simple, seemingly harmless requests (clicking a button, validating information, sending a text message) before progressing to more compromising requests. This gradual escalation reduces the perceived risk at each step.

Strengthening of Attentional Tunneling: The rapid sequence of actions under pressure keeps the victim in a state of intense focus, preventing the necessary detachment to reassess the situation.

Paradoxical Decrease in Vigilance: Completing several "protective" actions can generate a growing sense of security ("I am protecting my account"), further diminishing critical vigilance.

At this stage, the victim is in a state of "active compliance", characterized by increased trust in the interlocutor, a feeling of collaboration in their own protection, and a gradual reduction of initial anxiety. It becomes difficult for them to interrupt the engaged process. They may even feel a sense of relief at the end of the call.

Phase 4: Post-Fraud Awareness

This final phase of the model corresponds to the discovery of fraud and financial damage. It represents a major psychological shock. This moment can occur in different ways: a notification from the bank about suspicious transactions, direct observation of fraudulent movements on the account, or a gradual realization after hanging up, accompanied by a strange feeling and memories of inconsistencies in the attacker's speech.

Victims generally exhibit a combination of emotional and behavioral reactions (Foucault et al., 2025), ranging from anxiety related to financial consequences and invasion of privacy to shame for having been duped, especially among those who see themselves as vigilant and proficient with digital tools. Some also express feelings of guilt, self-blame ("How could I have been so naive?"), and anger towards the attacker and themselves.

As a result, several types of behavioral reactions can persist over time: compulsive account checks and communications, generalized mistrust towards institutional contacts, avoidance or, conversely, overinvestment in security measures, and reluctance to share the experience or file a complaint due to fear of social judgment.

This post-attack phase involves a cognitive reconstruction process, where the victim retrospectively reanalyses the event, identifying missed warning signals and moments of turning point. Although this reflexive analysis can be painful, it constitutes a potentially protective learning experience for the future.

THEORITICAL CONTRIBUTION OF THE CMAA MODEL

Our systemic model provides several significant contributions to the understanding of spear phishing. On one hand, the integrative approach highlights a harmful "cognitive cascade" where emotions play a central role. On the other hand, the model sheds light on the reality of risks faced by the public.

Unlike fragmented approaches that focus on isolated aspects (persuasion techniques, cognitive biases, or emotional responses), our model offers an integrated vision that links the temporal dimension across four phases, the strategic dimension of the attacker, the cognitive processes involved, and the emotional dimension of the victim. This holistic approach helps to understand how these different dimensions interact to create systemic vulnerability.

Our analysis reveals that the effectiveness of spear phishing relies on a cascade of cognitive degradations: each tipping point prepares and facilitates the next step by progressively impairing the victim's critical judgment. This sequential perspective is crucial for identifying the most effective points for preventive intervention.

The CMAA model also emphasizes the decisive role of emotions as vectors of manipulation. The transition from trust (phase 1) to fear (phase 2), and then to apparent relief (phase 3), constitutes a carefully orchestrated emotional journey that bypasses analytical reasoning and decision-making. It is important to note that this affective dimension is still too often underestimated in purely technical cybersecurity approaches.

Finally, our model offers a novel insight into a form of scam specifically targeting the public. To date, scientific literature has predominantly focused on cybersecurity issues in corporate settings, such as crisis management or employee resilience, largely overlooking the challenges faced by individuals (Damiano, 2023).

DISCUSSION AND RECOMMENDATIONS

This research aimed to answer a seemingly simple but fundamental question: why does human remain the weakest link in the face of spear phishing cyberattacks, despite increasing awareness of cyber risks?

Our study provides some answers by demonstrating that the formidable effectiveness of spear phishing does not result from mere "negligence" or "lack of vigilance" on the part of victims, but from a systematic and sophisticated exploitation of our fundamental cognitive mechanisms.

The holistic CMAA model, developed in four phases, reveals how attackers orchestrate a cascade of psychological manipulations. This detailed understanding of the underlying mechanisms allows us to go beyond the common observation ("90% of attacks succeed due to human error") and precisely identify where and how our cognitive processes are diverted.

Practical Implications

Our research findings suggest that a necessary paradigm shift is required in prevention and awareness approaches. Beyond technical innovations in engineering and design, our work advocates for a multidimensional cybersecurity approach where humans develop **situational awareness** and **protective metacognition.** It is no longer just about learning to recognize external warning signals (such as spelling mistakes or suspicious addresses), but about cultivating a sharp awareness of our own cognitive and emotional states during potentially risky interactions.

Specifically, this involves:

Training to recognize cognitive vulnerability states: learning to identify when we are under the influence of attentional tunneling, intense emotional responses, or confirmation biases (Aggarwal et al., 2025).

Developing "cognitive pause" protocols: establishing systematic mechanisms to step back during situations that exhibit the identified characteristics (urgency + authority + action request). These pauses can be initiated autonomously by the user or triggered by an automated notification on their phone.

Normalizing independent verification: creating a reflex which interrupts an interaction to verify through an alternative channel (for example, hanging up and calling back the bank) is not seen as mistrust but as a standard safety practice.

Destigmatizing victimization: understanding that falling into the spear phishing trap results from the exploitation of universal cognitive mechanisms, not personal weakness. This understanding can encourage sharing experiences and collective learning.

Our research, however, presents certain limitations that should be acknowledged.

Methodological limitations: Our study is based on a limited number of interviews, and these interviews rely on victims' retrospective accounts, which may introduce potential biases related to memory reconstruction. Victims might rationalize their behavior after the fact or omit certain details as a psychological defense mechanism.

Context-specificity: Our model was developed solely from cases of fake banking advisor fraud. Although the identified mechanisms are likely transferable to other forms of spear phishing (such as CEO scams, fake update requests, etc.), this generalization would require further empirical validation.

CONCLUSION

Beyond the scientific and practical contributions of this research, we would like to conclude this article with a proposal for a paradigm shift in how we approach digital security and cyber protection for the public.

Spear phishing reveals a fundamental paradox: as our technical systems become more complex, attackers increasingly exploit our own humanity, our emotions, trust, need for coherence, and survival instincts in the face of threats. In this sense, cybersecurity is not merely a technical issue but a matter of understanding the human condition.

The solution does not lie in eliminating our cognitive biases, which are often adaptive in other contexts, or in fostering a generalized mistrust that would paralyze our social interactions. Instead, it resides in developing a form of "digital clairvoyance": the ability to recognize contexts where our cognitive automatic responses can be exploited, and to consciously activate appropriate protective strategies.

This perspective repositions humans not as the "weak link" to be fixed, but as conscious and reflective actors capable, through understanding their own mental mechanisms, of developing genuine resilience against digital manipulations.

In a world where social engineering techniques will continue to evolve in sophistication, especially with the increasing capabilities of generative artificial intelligence, this situational intelligence and cybersecurity metacognition may be our best defense. Not to become invulnerable, but to significantly reduce our exposure and to turn each attack attempt into an opportunity for learning, ultimately serving our own protection.

ACKNOWLEDGMENT

The authors would like to acknowledge the multidisciplinary team "Digital Protection Services" at Orange Innovation in Lannion.

REFERENCES

Aggarwal, Anu & Ferreira, Maria & Aggarwal, Palvi & Rajivan, Prashanth & Gonzalez, Cleotilde. (2025). Cognitive Biases in Cyber Attacker Decision Making: Translating Behavioral Insights Into Cybersecurity.

Cialdini, R. B. (2021). Influence and Manipulation: The Psychology of Persuasion. Translated from English by E. Debon and F. Paban.

Damiano, Jean-Pierre. (2023). Cybersecurity: context, challenges, current situation, and outlook. https://www.researchgate.net/publication/368308456_La_cybersecurite_contexte_enjeux_constats_et_perspectives_Cybersecurity_context_issues_findings_and_outlook

Demarchi, Samuel & Py, Jacques. (2006). The cognitive interview: its effectiveness, application, and specific characteristics. Revue Québécoise de Psychologie. 27. 177–196.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66(3), 183–201.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford University Press.

Foucault, B., Salmon, N., & Coly, A. (2025). It all happened so fast! Victims of online fraud share their stories. https://hellofuture.orange.com/fr/tout-est-alle-tres-vite-les-victimes-de-fraudes-en-ligne-nous-racontent/

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. Journal of Personality and Social Psychology, 65(2), 221–233.
GSMA (2023). White Paper: Improving CLI Validity – Solutions and Regulatory Assessment. Version 1.0.
Guéguen, N. (2010). Chapter 1. The Milgram paradigm: being compelled to violence through simple obedience. Authority and submission (pp. 5–37). Dunod. https://shs.cairn.info/autorite-et-soumission--9782100550470-page-5?lang=fr
Johnson, Chelsea & Gutzwiller, Robert & Gervais, Joseph & Ferguson-Walter, Kimberly. (2021). Decision-Making Biases and Cyber Attackers.
Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.
Kaspersky (2025). https://www.kaspersky.fr/resource-center/definitions/what-is-social-engineering
Krombholz, Katharina & Hobel, Heidelinde & Donko-Huber, Markus & Weippl, Edgar. (2014). Advanced social engineering attacks. Journal of Information Security and Applications. 22. 10.1016/j.jisa.2014.09.005.
LeDoux, J. E. (1996). The emotional brain: The mysterious underpinnings of emotional life. Simon & Schuster.
Livet, P. (2016). Vigilance and Negligence. In: Intellectica. Revue de l'Association pour la Recherche Cognitive, n°66, 2016/2.
Longtchi, Theodore & Rodriguez, Rosana & Al-Shawaf, Laith & Atyabi, Adham & Xu, Shouhuai. (2024). Internet-Based Social Engineering Psychology, Attacks, and Defenses: A Survey. Proceedings of the IEEE.
Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology, 67(4), 371–378.
OpinionWay pour Orange (2025). Study on "fake bank advisor" fraud. Septembre 2025.
Paillé, P. (2017). Chapter 3. Grounded theory analysis. In M. Santiago Delefosse and M. Del Rio Carral, *Qualitative Methods in Psychology and Health Sciences*. (pp. 61–83). Dunod. https://doi.org/10.3917/dunod.santi.2017.01.0061