Le tournant cognitif de la cybersécurité : changement de paradigme et prolégomènes à la cybersécurité cognitive.

Bruno Teboul

Le tournant cognitif de la cybersécurité : changement de paradigme et prolégomènes à la cybersécurité cognitive.

Bruno Teboul

Paris, le 27 mars 2022.

Introduction

En janvier, le Forum Economique Mondial a publié son rapport sur les risques mondiaux pour 2022[1]. Selon le rapport, 95% des problèmes de cybersécurité sont dus à des erreurs humaines. Et le volume d'e-mails malveillants auquel nous sommes confrontés en tant qu'utilisateurs ne cesse d'augmenter : en un an seulement, le volume des « malwares » a augmenté de 358% et celui des « ransomwares » de 435%. Cette augmentation du nombre de cyberattaques a des conséquences désastreuses sur l'économie : 6 000 milliards de dollars de coûts cumulés. On constate que 94% des cyberattaques ont été déclenchées par la diffusion d'un email malveillant. Ce constat effroyable est sans appel et doit nous interroger sur les causes d'une telle réussite de la part des cybercriminels.

En effet, comment expliquer cette prolifération de la criminalité numérique à l'heure où l'intelligence artificielle apparait comme le seul rempart efficace face aux cyberattaques les plus sophistiquées ?

« *L'intelligence artificielle a été utilisée dès le début des années 2010 pour protéger les systèmes d'information grâce à des composantes d'apprentissage automatique incorporées dans des solutions de supervision des réseaux. Après cette phase initiale d'entrainement, les composantes sont en mesure de détecter des séquences d'évènements en déviation par rapport à la normalité apprise puis de produire des alertes d'attaques probables. L'apprentissage statistique permet ainsi de détecter des menaces inédites qui ne figurent pas dans les bases historiques d'attaques connues contrairement aux antivirus classiques qui s'appuient sur des bases de signatures de virus et de malwares connues. Cette approche data centrée permet aujourd'hui de construire des solutions efficaces de SIEM (Security Information and Event Management) orientées UEBA (User and Entity Behavior Analytics)* »[2].

Alors, pourquoi les solutions actuelles d'analyse et de diagnostic automatiques des failles de sécurité numérique sont-elles malgré tout mises en échec par les pirates informatiques ?

1 https://www3.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2022.pdf

[2] Bruno Teboul et Thierry Berthier, LES GRANDS DOSSIERS DE DIPLOMATIE N° 66 **:** « Les usages malveillants de l'intelligence artificielle au service de la cybercriminalité », Géopolitique de la criminalité**,** Areion Group, Février- Mars 2022.

L'explication causale serait-elle véritablement « technologique » ? Quel rôle et quelle place le facteur humain occupe-t-il réellement dans ce processus ? Comment lutter contre ce fléau numérique dont l'origine serait bien plus humaine que technologique ?

1/ Les limites de l'intelligence artificielle face aux cyberattaques par ingénierie sociale.

Dans un article récent nous avions dressé un panorama non exhaustif des détournements potentiels de l'apprentissage automatique mettant en lumière toute l'étendue et la puissance de ces technologies lorsqu'elles sont employées à des fins malveillantes. Nous avions alors distingué « *deux grandes familles de détournements malveillants fondés sur l'utilisation de l'intelligence artificielle : la première concerne les actions malveillantes dirigées par l'IA dans le cyberespace, la seconde réunit les opérations malveillantes ou criminelles opérées dans l'espace physique, supervisées ou dirigées par l'IA. Dans les deux cas, les dommages et l'impact sur la cible peuvent être de très haut niveau tout en préservant la sécurité et l'anonymat de l'attaquant. Dans chacun des cas, l'intelligence artificielle agit comme un facteur multiplicateur de puissance ou comme un outil de saturation de l'espace des attaques pour une cible non préparée à ces nouvelles menaces* »[3].

Pour rappel, nous nous concentrons ici sur les cyberattaques par ingénierie sociale, que nous avons définit comme « *des attaques informatiques qui exploitent les failles et les faiblesses psychologiques, bien humaines, en tentant de persuader un individu (une victime) à agir comme prévu, selon un scénario malicieux et efficace à la fois. Ces attaques informatiques exploitent les faiblesses des interactions humaines et des constructions comportementales/culturelles qui se produisent sous de nombreuses formes, y compris le « phishing », « l'escroquerie », les « fraudes », les « spams », le « spear phishing » et les « sock puppets » sur les réseaux sociaux »[4].*

Dans l'article précité nous avions clairement indiqué que la cause majeure à ce problème ne trouve pas sa source dans les failles et les insuffisantes des technologies en œuvre. Les technologies les plus avancées en matière de cybersécurité utilisent des algorithmes d'intelligence artificielle éprouvés et reposant sur l'apprentissage statistique relativement efficaces lorsqu'ils sont intégrés dans des solutions logicielles de type « SIEM UEBA » pour contrer les menaces connues en matière d'attaques sur les réseaux notamment.

Elles permettent de bloquer nombre d'attaques comme les compromissions d'emails notamment : « *en 2017, lors de la conférence BlackHat USA, une équipe de chercheurs a montré qu'il était possible d'utiliser les techniques de Machine Learning afin d'analyser des données liées aux attaques de compromission d'e-mails commerciaux. Cette forme de cybercriminalité utilise la fraude par e-mail pour escroquer les organisations et identifier les cibles potentielles d'attaques. Le système développé exploite à la fois les fuites de données et les informations librement disponibles sur les réseaux sociaux. Notamment, sur la base de son historique, le système peut prédire avec précision si une attaque réussira ou non* ».

---

[3] Bruno Teboul et Thierry Berthier, LES GRANDS DOSSIERS DE DIPLOMATIE N° 66 **:** « Les usages malveillants de l'intelligence artificielle au service de la cybercriminalité », Géopolitique de la criminalité**,** Areion Group, Février- Mars 2022.

[4] Bruno Teboul, « Approche cognitive des cyberattaques par ingénierie sociale ». *The European Scientist*, 2021. ⟨hal-03325684⟩.

L'intelligence artificielle est également efficace contre le camouflage et de l'encapsulation automatique de logiciels malveillants (« malwares »), tout comme sur la pénétration et la violation des bases de données, ou la compromission de services cloud ou d'assistants intelligents piégés, ou encore pour stopper la compromission automatique des mots de passe ou des CAPTCHA, et parfois même pour résister contre l'usurpation d'identité par « *escalade de privilèges* ».

Mais les cyberattaques qui nous préoccupent ici, sont les cyberattaques par ingénierie sociale qui agissent comme des attaques psychologiques en profitant et en jouant sur les vulnérabilités cognitives des utilisateurs. Nous parlons bien de cyberattaques qui misent sur les faiblesses de l'âme humaine, « *humaine trop humaine* » sans doute et pour lesquelles l'intelligence artificielle n'est d'aucun usage ou d'aucune efficience démontrée. Comment expliquer qu'un individu puisse cliquer sur un mail malveillant et ouvrir une pièce jointe corrompue, « vérolée » ?

Si toutes ces attaques étaient déjouées par des algorithmes nous n'aurions pas à déplorer cette croissance exponentielle des cyberattaques. La vraie raison de cet échec monumental et délétère sur les organisations publiques et privées de tous secteurs et de toutes tailles s'explique par une mauvaise compréhension du problème, une erreur de diagnostic fatal, létal.

« *Le succès des cyberattaques d'ingénierie sociale est inversement lié à leur prévalence, cela pose ainsi un dilemme : lorsque les défenses automatisées sont efficaces pour détecter et filtrer la plupart des cyberattaques d'ingénierie sociale, les attaques restantes ont plus de chances d'aboutir. Une approche pour faire face à ce dilemme est de recourir aux principes de la psychologie cognitive. On sait que la majeure partie du traitement de l'information neuronal est isolé de la conscience. Certaines informations pourraient être consciemment utilisées, mais elles peuvent aussi ne pas être conscientes à un moment donné. Notre système visuel, par exemple, calcule la profondeur 3-D à partir d'entrées rétiniennes 2-D. Nous n'expérimentons pas consciemment les calculs nécessaires pour transformer l'entrée 2-D en une perception 3-D. Au lieu de cela, nous sommes conscients du résultat produit (c'est-à-dire de voir un monde en 3D) mais pas du processus qui a conduit au même résultat final. Les influences du traitement subconscient sont bien connues pour avoir un impact sur le comportement humain* »[5].

2/ L'erreur de raisonnement et de diagnostic à l'origine de la prolifération des cyberattaques par ingénierie sociale.

La cybersécurité ne peut pas faire l'impasse sur le raisonnement scientifique : comprendre les causes profondes des cyberattaques par ingénierie sociale nous impose la même rigueur que la science médicale, où le but du raisonnement est de découvrir par l'examen de ce que l'on sait déjà quelque autre chose qu'on ne sait pas encore : *« les trois étapes qui scandent le raisonnement médical sont le diagnostic, le pronostic et la décision thérapeutique. Le diagnostic est à l'évidence l'étape la plus importante puisqu'elle conditionne les deux autres.*

[5] Bruno Teboul. THE EUROPEAN SCIENTIST Approche cognitive des cyberattaques par ingénierie sociale. *The European Scientist*, 2021. ⟨hal-03325684⟩.

Sans identification précise de la maladie en cause, il est illusoire de prétendre établir un pronostic et a fortiori engager une action thérapeutique. Le diagnostic est à la démarche médicale ce que l'identification d'un coupable est pour la justice. Le pronostic est une projection dans l'avenir. À ce titre, il est une conjecture qui doit prendre en compte de nombreux facteurs : l'évolution naturelle de la maladie, son évolution sous traitement mais également des facteurs propres au patient. Ainsi, pour une même affection, le pronostic ne sera pas le même pour tous les patients »[6].

A ce jour, les sociétés de cybersécurité qui tentent de comprendre et de décrypter les causes des cyberattaques par ingénierie sociale font fausse route, tout en évoquant l'origine « humaine » du problème, elles font une erreur de diagnostic et de thérapeutique : « *nos chercheurs spécialisés dans la cybermenace constatent que les attaques sont de plus en plus ciblées, c'est-à-dire individualisées. Le principe de l'arrosoir a fait son temps. Une fois qu'une entreprise est dans le collimateur d'un groupe de hackers, l'attaque est souvent minutieusement planifiée. Les médias sociaux sont passés au crible, les collaborateurs et leurs responsabilités ainsi que les possibles autorisations qui en résultent sont évalués. Des organigrammes sont créés et souvent, les échanges de mails sont lus, sans intervention active, juste pour comprendre la culture et le langage de l'entreprise. Lorsque toutes ces pièces du puzzle sont réunies, le collaborateur – souvent peu méfiant – n'est plus qu'à un clic d'une cyberattaque réussie »*[7].

Notre propos n'est pas de désigner ici, telle ou telle entreprise du secteur cybersécurité (cette interview d'un dirigeant de Proofpoint montre qu'ils sont très performants et conscients des enjeux) mais plutôt d'illustrer les raisonnements à l'œuvre, qui certes, pointent bien le problème fondamental de la vulnérabilité du l'utilisateur, mais ignorent totalement l'analyse psychologique et cognitive qui en est la cause profonde.

Il n'y a rien d'étonnant malheureusement, nous avions déjà précisé que les études scientifiques actuelles se concentraient sur l'analyse des failles techniques et ignoraient totalement la causalité cognitive de ces attaques. Par exemple, Gupta et al. (2016) enquêtent sur les défenses technologiques contre les attaques de phishing.

Salahdine et Kaabouch (2019) passent en revue les cyberattaques d'ingénierie sociale et les stratégies d'atténuation, mais ils ne discutent pas de facteurs tels que la cognition humaine. Darwich et al. (2012) analysent la relation entre les facteurs humains, les cyberattaques d'ingénierie sociale et les stratégies de cyberdéfenses, mais ils n'examinent pas ce qui rend un individu psychologiquement vulnérable aux cyberattaques d'ingénierie sociale[8].

« L'attaque externe n'est pas le seul scénario dans lequel le comportement humain fait la différence. Dans le cas des insider threats, le danger ne vient pas de l'extérieur, mais se trouve

---

[6] Masquelet, Alain-Charles. « Les grandes étapes de la démarche médicale », Alain-Charles Masquelet éd., *Le raisonnement médical.* Presses Universitaires de France, 2006, pp. 7-16.

[7] L'humain au cœur de la cybersécurité, interview de Irène Marx, Country Manager Suisse chez Proofpoint.

[8] Bruno Teboul. THE EUROPEAN SCIENTIST Approche cognitive des cyberattaques par ingénierie sociale. *The European Scientist*, 2021. ⟨hal-03325684⟩.

au cœur de l'entreprise. Il convient de distinguer deux types d'insiders : le collaborateur naïf qui agit en toute bonne foi et révèle des données de l'entreprise sans le savoir – un risque particulièrement élevé en télétravail. Et les collaborateurs qui agissent sciemment et intentionnellement contre leur propre entreprise. Dans les deux cas, les données de l'entreprise tombent entre les mains de tiers. Et à chaque fois, c'est le comportement humain qui explique le nombre élevé d'incidents »[9].

La « naïveté » de certains collaborateurs ou bien leur éventuelle « corruption » et/ou leur intention délibérée de nuire à leur entreprise, ne sont pas des causes suffisantes pour expliquer les mécanismes mis en place par les hackeurs. Il faut comprendre clairement et distinctement la nature « psychologique » des pièges cognitifs tendus par ces pirates informatiques, afin de les diagnostiquer et de mieux les prévenir. Cette compréhension des causes psychologiques est cruciale, car elle va permettre de savoir pourquoi et comment un salarié est amené à prendre la décision de cliquer ou non sur un mail malveillant…

Tous les salariés sont des êtres humains, confrontés à leurs propres faiblesses psychologiques, dans leur prise de décision et à leurs états psychologiques, ils sont avant tout victimes de leur état psychologique et mental avant et pendant l'attaque : fort niveau de stress, baisse du seuil de vigilance, charge cognitive trop élevée, émotions négatives...

Tout être humain est victime de *biais cognitifs* en tout genre, il procède par heuristique ou justifie souvent ces choix *a posteriori* : ce qui plaide pour une remise en cause profonde du principe même de rationalité. C'est le principe cartésien de rationalité qui a « volé en éclat » sous la pression de l'observation par imagerie médicale (en 1994 avec les publications de A. Damasio notamment) et par les avancées de la psychologie du raisonnement (notamment les travaux de Kahneman & Tversky dès 1974). C'est pourquoi nous pouvons parler de rupture épistémologique, au sens de rupture de paradigme, véritable rupture théorique et pratique qui relativise un certain héritage cartésien réfuté par Damasio qui n'hésite pas à parler dans l'un de ses ouvrages de référence de « *l'erreur de Descartes* ».

En effet, nous savons désormais que la prise de décision peut être conduite de deux manières fondamentalement différentes. La première se produit via des processus relativement automatiques qui sont rapides mais qui peuvent ne pas être un choix optimal dans certains cas, appelés « heuristiques » et « biais » (Tversky et Kahneman, 1974 ; Gigerenzer, 2008).

La deuxième approche consiste à utiliser un raisonnement de traitement conscient et contrôlé, qui est plus lent mais qui peut être plus sensible aux particularités d'une situation donnée. Le stress a une variété d'effets sur la prise de décision et de nombreuses subtilités (Starcke et Brand, 2012), mais il peut, en général, altérer la prise de décision rationnelle.

Ces chercheurs en économie dite expérimentale ou « behavioural economics » que sont Daniel Kahneman et Amos Tversky ont développé une théorie sur les « heuristiques et les biais dans le jugement » et font l'hypothèse que les erreurs systématiques par rapport à la norme bayésienne, et mises en évidence par leurs expériences, s'expliquent par des raccourcis mentaux, plus faciles d'accès, moins coûteux en temps et en concentration ou heuristique.

[9] L'humain au cœur de la cybersécurité, interview de Irène Marx, Country Manager Suisse chez Proofpoint.

Quand le coût du calcul bayésien normatif est important, l'individu a tendance à faire appel à des exemples qui l'ont marqué ou à des aspects parlants du problème qui lui permettent de réaliser des approximations. De ce fait, la décision se fait via un arbitrage entre ces raccourcis et un calcul plus rigoureux.

On sait que le stress peut influencer la prise de décision (non-rationnelle, non bayésienne), en altérant l'attention, une composante vitale de la mémoire de travail, de manière aussi bien bénéfique que préjudiciable (Al'Absi et al., 2002). L'effet tunnel attentionnel est l'un de ces effets du stress aigu où l'attention est « *hyper-focalisée* » sur les aspects pertinents à la cause du stress, mais elle est moins sensible aux autres informations.

Le terme « *tunneling* » dérive de l'utilisation de tâches d'attention spatiale, où l'excitation due au stress conduit les sujets à ignorer des choses qui sont plus éloignées de leur centre d'attention (Mather et Sutherland, 2011).

Une étude du Cesin et du cabinet Advens, a récemment montré que les responsables de la cybersécurité eux-mêmes subissent un niveau de stress préoccupant dû à la crise sanitaire (août 2021)[10] : 60 % des responsables de la sécurité informatique (RSSI) sont en situation de stress élevé. « *Depuis la crise sanitaire, les entreprises subissent une recrudescence des cyberattaques. En première ligne, les responsables de la sécurité et de l'informatique éprouvent un niveau de stress particulièrement élevé. En témoigne le Club des Experts de la Sécurité de l'Information et du Numérique (Cesin) qui vient de révéler les chiffres d'une enquête inédite sur l'état psychologique des responsables cyber* ».

Nous le savons, appliqué à la cybersécurité, le « *tunneling d'attention* » à partir d'un message de phishing peut conduire à une hyper-concentration sur le texte de l'e-mail mais peut conduire à ignorer une adresse suspecte ou à passer à côté d'alertes périphériques. La mémoire de travail est également vulnérable au stress (Schwabe et Wolf, 2013), notamment en interférant avec la fonction du cortex préfrontal (Elzinga et Roelofs, 2005 ; Arnsten, 2009).

3/ Le tournant cognitif de la cybersécurité : prolégomènes à toute cybersécurité future qui se présentera comme science...

Nous nous proposons d'établir les prolégomènes à cette nouvelle discipline scientifique, qui émerge dans le champ de la cybersécurité : la « *cybersécurité cognitive* » ou pour introduire un néologisme, la « *neuro-cybersécurité* ». La cybersécurité cognitive ou neuro-cybersécurité constitue un véritable « *changement de paradigme* », au sens de Polanyi[11] et de Kuhn[12].

[10] https://www.cesin.fr/actu-le-cesin-et-advens-revelent-les-resultats-dune-grande-enquete-inedite-sur-le-stress-des-responsables-cyber.html

[11] Michael Polanyi, Science, faith and society, Chicago, University of Chicago Press, 1964.

[12] Thomas Kuhn (trad. de l'anglais par Laure Meyer), *La structure des révolutions scientifiques* [« The Structure of Scientific Revolutions »], Paris, Flammarion coll. « Champs / 791 », 2008 (1re éd. 1962).

Un changement de paradigme est un changement majeur dans les concepts et les pratiques de la façon dont quelque chose fonctionne ou est accompli. Un changement de paradigme se produit très souvent lorsqu'une nouvelle technologie est introduite qui modifie radicalement le processus de production d'un bien ou d'un service. Les paradigmes sont importants car ils définissent la façon dont nous percevons la réalité. En tant que tel, chacun est soumis aux limitations et distorsions produites par sa nature socialement conditionnée. Ces changements sont devenus beaucoup plus fréquents au cours des cent dernières années, alors que la révolution industrielle a transformé de nombreux processus sociaux et industriels. Ces processus sont susceptibles de devenir encore plus courants à l'avenir à mesure que notre taux d'avancement technologique augmente.

Les changements de paradigme se produisent dans un large éventail d'autres contextes pour décrire un changement profond de modèle, de modalité ou de perception. Ceux que l'on trouve dans le monde scientifique résultent souvent de scientifiques travaillant en marge. Leurs recherches controversées sont perçues comme malavisées ou comme une impasse.

Alors que le scepticisme et l'enquête font partie intégrante du processus scientifique, un scientifique a parfois une révélation, ce qui conduit à un changement de paradigme. Le poids de la résistance scientifique et publique au nouveau paradigme peut parfois provoquer le ridicule. Bien qu'elle ne soit pas instantanément acceptée, s'il est prouvé qu'une science marginale repose sur des bases solides, l'élan se construit lentement contre le paradigme établi. Tel est le cas de la cybersécurité, en tant que discipline technoscientifique, crantée, figée au stade « technologique » et qui pense son salut à l'aune de l'automatisation algorithmique et de l'intelligence artificielle. Mais le cerveau humain résiste, persiste et signe le dépassement des remèdes technologiques limités et contre-productifs, face aux cyberattaques.

Un vrai changement de paradigme s'opère par la réorientation de l'analyse des failles et des vulnérabilités « *techno-centrées* », vers le prisme d'une analyse « neurocognitive » : où les faiblesses de l'esprit humain, sont démontrées, diagnostiquées et mises à l'épreuve par des tests psychologiques qui nous indiquent, pourquoi et comment les humains sont vulnérables et peuvent à tout comment cliquer sur un lien malveillant et déclencher une catastrophe en chaine.

Quel est ce tournant paradigmatique et de quelle rupture épistémologique parle-t-on pour la cybersécurité ?

Les découvertes scientifiques des neurosciences se confirment, et la cybersécurité ne peut plus l'ignorer désormais : les conséquences doivent faire évoluer les analyses des spécialistes en cyber à la lumière de ces nouvelles connaissances qui fondent une véritable « rupture épistémologique » au sens de Bachelard et imposent à la discipline d'évoluer et de s'adapter en théorie comme en pratique.

On peut résumer les travaux de Bachelard sur l'histoire des sciences par une thèse devenue aujourd'hui familière : la discontinuité. Le progrès, pour Bachelard, est incontestable, il constitue la dynamique même de la culture scientifique. *« Pour la pensée scientifique, le progrès est démontré, il est démontrable»*. Mais il ne s'effectue pas selon une marche

régulière et ininterrompue. L'histoire des sciences n'est pas une simple accumulation de découvertes et d'inventions qui s'additionneraient progressivement, mais une aventure faite de perpétuelles ruptures.

D'où le premier concept, celui de rupture épistémologique, correspondant à ces mutations brusques qui apportent des impulsions inattendues dans le cours du développement scientifique, dans la construction d'une discipline scientifique, mais cela vaut aussi pour une technoscience comme la cybersécurité qui n'a cessé d'évoluer au fil du temps.

Parce-que l'origine du problème des cyberattaques par ingénierie sociale (le diagnostic) est de nature neurocognitive et non pas technique et la solution pour lutter efficacement (la thérapeutique) contre cette maladie numérique très contagieuse et délétère réside dans la mise en place urgente d'une analyse des facteurs cognitifs en jeu, mais également par la définition d'un nouveau cadre théorique et méthodologique appropriés et appliqués à la cybersécurité.

Dans le cadre de notre approche théorique, nous devons associer l'analyse de certains biais cognitifs connus de la littérature et des hackeurs, à des traits de personnalité, qui traduisent un état psychologique défini comme le stress, la vigilance, la charge cognitive…

Voici quelques exemples de biais cognitifs couramment utilisés par les hackeurs, pour piéger psychologiquement les victimes de cyberattaques par ingénierie sociale :

- **La tentation de l'appât du gain (l'avidité) et son rôle dans la vulnérabilité aux cyberattaques.**

Dans un email de type « phishing », où les utilisateurs sont invités à cliquer pour remporter une importante (jeux en ligne, loteries…), les utilisateurs se font piéger par leurs désirs et leurs émotions liés à la tentation de gagner facilement de l'argent, par l'appât du gain (ou bien « avidité »). En effet, de nombreuses études ont montré que l'obtention d'un gain active deux principales régions cérébrales : le centre des émotions du cerveau (l'amygdale cérébrale) et le circuit de récompense (notamment le noyau accumbens), responsable de la sensation de plaisir.

Les études menées notamment par les psychologues Brian Knutson et Lisbeth Nielsen montrent que les jeunes adultes ont tendance à penser que le gain obtenu sera encore plus stimulant que son anticipation. Ces derniers sous-estiment l'excitation que leur procure la perspective d'un gain d'argent.

En revanche, avec le temps et l'expérience, les adultes alors prennent conscience que la perspective du gain a un effet au moins aussi puissant que sa réalisation. Et pourtant, malgré cette prise de conscience, l'envie de gagner toujours plus persiste ! C'est pourquoi ce type de cyberattaque fonctionne sur de nombreux utilisateurs et les hackeurs le savent bien.

- **Le rôle de la curiosité (indiscrétion) dans la vulnérabilité aux cyberattaques.**

Certains utilisateurs peuvent recevoir des mails malveillants qui leur propose de se connecter à un service de partage de fichiers, afin de découvrir les bulletins de salaires de leur entreprise. Ils sont ainsi incités à se connecter dans un délai limité en général très court (48 heures) ce qui renforce et conditionne une prise de décision rapide et attise davantage la curiosité.

La curiosité est une attitude complexe et son objet, c'est l'information : les personnes curieuses veulent savoir ce qu'elles ne savent pas encore (logique). La littérature scientifique confirme cette affirmation, selon George Loewenstein la curiosité apparaît lorsque l'on réalise qu'il y a une « absence », un « trou », dans notre savoir. Ce manque créerait un sentiment de déficience nous motivant à combler ce que l'on ne sait pas.

Marret Noordewier et Eric van Dijk, se sont intéressés à la manière dont nous « vivons » la curiosité. Selon eux, elle peut créer des émotions plus ou moins plaisantes. Selon Marret Noordewier et Eric van Dijk, ces émotions pourraient être influencées par un facteur : le temps. Les deux scientifiques ont mené trois études, avec plus de 200 participants, pour comprendre le phénomène.

Les résultats de ces recherches, publiés dans la revue Cognition and Emotion, suggèrent que lorsque notre curiosité ne peut pas être satisfaite immédiatement, nous aurions tendance à nous concentrer sur le fait qu'on ne sait pas, sur notre frustration…

Ce qui serait, logiquement, source d'émotions négatives. À l'inverse, lorsqu'on sait que notre curiosité va bientôt être assouvie, nous pourrions alors centrer notre attention sur la résolution future de nos préoccupations. Cela permettrait une expérience beaucoup plus positive de la curiosité.

C'est exactement ce qui se produit quand on reçoit un email « malveillant » avec la possibilité de découvrir les salaires de ses collègues. Le désir de savoir combien gagne les autres salariés de son entreprise est certes indiscret, contestable et sans doute illégitime, mais la curiosité exacerbée ne peut être qu'assouvie et finit toujours par l'emporter sur le comportement (non-rationnel) de l'utilisateur…

- **Le rôle du stress dans la vulnérabilité aux cyberattaques**

Parfois les hackeurs utilisent des stratagèmes efficaces pour envoyer des courriels piégés dans lesquels, il sollicite la dimension psychologique relative au stress : il est souvent demandé aux utilisateurs de mettre à jour leurs données de connexion sous 24 ou 48 heures, sans quoi l'accès aux services informatiques de la société sera suspendu.

Le stress influence la cognition et les comportements des victimes de cyberattaques. Il faut distinguer le stress aigu du stress chronique, le stress chronique commençant après une durée de l'ordre de quelques mois, car son impact sur la cognition peut différer et le stress chronique est plutôt catégorisé comme un facteur d'influence sur le long terme. Les réponses neurobiologiques et hormonales à un événement stressant ont été relativement bien

analysées dans la littérature spécialisée, tout comme leur impact sur le comportement (Lupien et al., 2009).

L'effet tunnel attentionnel est l'un de ces effets du stress aigu où l'attention est hyper-focalisée sur les aspects pertinents à la cause du stress, mais elle est moins sensible aux autres informations. Le terme tunneling dérive de l'utilisation de tâches d'attention spatiale, où l'excitation due au stress conduit les sujets à ignorer des choses qui sont plus éloignées de leur centre d'attention (Mather et Sutherland, 2011).

C'est exactement ce qui se produit lorsqu'un utilisateur reçoit ce type de mail, le tunneling d'attention le conduit à une hyper-concentration sur le texte de l'e-mail mais peut conduire à ignorer une adresse suspecte ou à passer à côté d'alertes périphériques.

Pour les psychologues, la « personnalité » est un terme technique qui diffère quelque peu de l'usage ordinaire. Il fait référence aux différences individuelles affectant les pensées, les sentiments et les comportements qui sont relativement cohérents au fil du temps et des situations. Nous disons « relativement » parce que, comme indiqué ci-dessus, les pensées, les sentiments et les comportements dépendent fortement de la situation, et les approches axées sur la durée de vie ont prouvé des changements notables de la personnalité avec l'âge (Donnellan et Robins, 2009).

- **Un nouveau cadre méthodologique : notre « *Neurocyber Framework* ».**

Les recherches en psychologie cognitive sur la personnalité sont dominées par le « **Framework Big 5** » de la personnalité. Ce cadre théorique et méthodologique a été développé durant une grande partie du vingtième siècle sous diverses formes (Digman, 1997). Le Framework Big 5 est basé sur des méthodes statistiques (analyse factorielle) qui identifient des dimensions abstraites qui peuvent économiquement expliquer une grande partie de la variance dans les mesures de la personnalité.

Les facteurs sont analysés sont : conscience, amabilité, neuroticisme, ouverture à l'expérience et extraversion. De nombreuses études sur la relation entre l'ingénierie sociale et la personnalité se concentrent sur l'ouverture, la conscience et le neuroticisme, qui sont considérés comme ayant le plus d'impact sur la vulnérabilité et la sensibilité aux attaques par ingénierie sociale.

Les facteurs qui composent le Framework Big 5 sont les suivants :

1. Ouverture : la volonté d'expérimenter de nouvelles choses.
2. Conscience : favorise les normes, fait preuve de maîtrise de soi et l'autodiscipline et la compétence.
3. Extraversion : être plus convivial, extraverti et interactif avec plus de monde.
4. Agréabilité : être coopératif, désireux d'aider les autres et croire à la réciprocité.
5. Neuroticisme : tendance à ressentir des sentiments négatifs, de la culpabilité, le dégoût, la colère, la peur et la tristesse.

L'expertise est généralement limitée à un domaine relativement étroit et ne se transfère pas à d'autres domaines (aussi facilement que nous aurions tendance à le croire). Ce point théorique est aussi appelé « *problème de transfert* » dans la littérature (Kimball et Holyoak, 2000). Le transfert limité d'expertise peut être aggravé par des illusions cognitives telles que l'effet Dunning-Kruger.

L'effet Dunning-Kruger montre empiriquement que les individus surestiment souvent leur compétence par rapport à leur performance objective (Kruger et Dunning, 1999). De même, « *l'illusion du savoir* » montre que les gens en savent généralement beaucoup moins sur un sujet qu'ils ne le croient, comme le révèle l'article de Keil publié en 2003.

Dans le domaine de la cybersécurité, ces phénomènes empiriques renforcent la confiance des utilisateurs. Une expertise restreinte en matière de cybersécurité peut être bénéfique, mais une expertise informatique très générale ne suffit pas à conférer des avantages ou des bénéfices réels en matière de sécurité.

Le Framework Big 5 nous parait solide et validé pour une approche expérimentale en France, en revanche, il est incomplet quant au nombre et à la nature des facteurs à évaluer dans le cadre de notre démarche en psychologie cognitive appliquée à la cybersécurité.

Dans la littérature, il est possible de distinguer deux approches théoriques qui conduisent à deux catégories de tests psychométriques : les types, comme par exemple le MBTI et les traits comme dans le cadre du Big 5.

La théorie des types met en relief des contrastes sous forme de profil, de modèle ou d'archétype. Ces profils sont formés par un ensemble de caractéristiques personnelles. Le MBTI est de cette catégorie. Il définit 16 types psychologiques décrits à partir de quatre dimensions : attitudes, fonctions de perception, fonctions de jugement et styles de vie.

Le test **Myers-Briggs Test Indicator (MBTI)** a été construit à partir des travaux de Carl G. Jung sur la théorie des types psychologiques. Il évalue quatre dimensions, dont trois qui avaient été définies par Jung. La première est l'attitude-type : elle est introvertie ou extravertie. La deuxième est la fonction de perception préférée par la personne : soit la sensation ou l'intuition. C'est le mode préféré des individus en ce qui a trait à la prise d'information. La troisième est la fonction de jugement ou de raisonnement préférée : soit la pensée ou le sentiment. À ces trois dimensions qu'avait déterminées Jung, Myers et Briggs en ont ajouté une quatrième : le style de vie. Il est soit du style « Jugement » ou du style « Perception ».

Le MBTI est un test de préférences qui met en relief quatre dichotomies : extraversion (E) et introversion (I), sensation (S) et intuition (N), pensée (T) et sentiment (F), ainsi que jugement (J) et perception (P). Plusieurs changements ont été apportés au MBTI depuis sa première édition. Parmi ceux-ci, les femmes et les hommes sont maintenant évalués de la même façon.

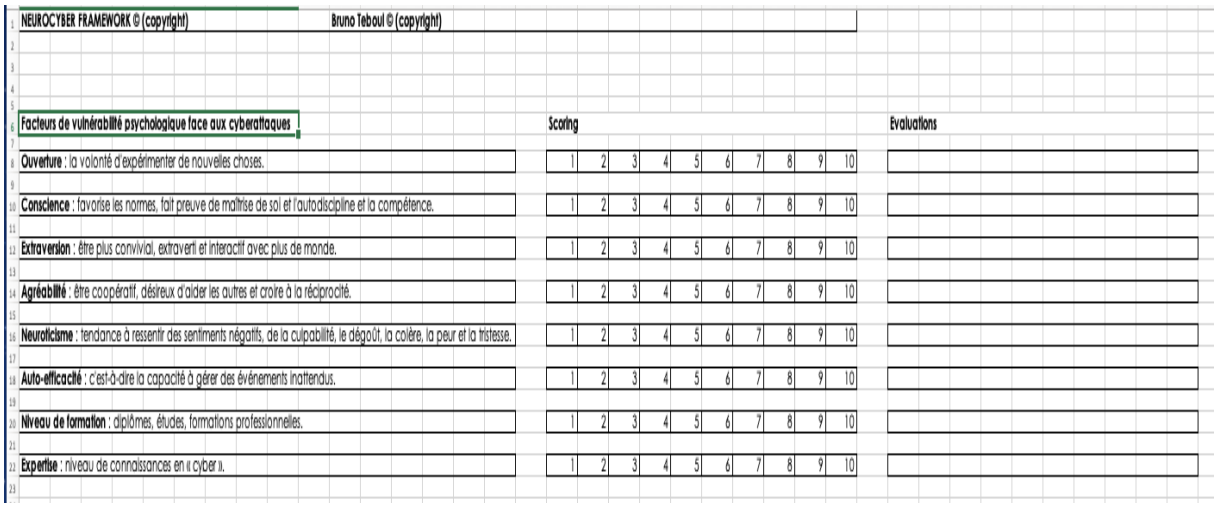
En nous inspirant à la fois du Big 5 mais en le complétant de 3 critères supplémentaires (voir tableau infra), nous pourrions automatiser les questionnaires, pondérer les scores et les combiner pour définir des types (dans l'esprit du MBTI), mais appliqué à l'évaluation du risque

cyber individuel et donc profiler des types de personnalité plus ou moins vulnérables face aux cyberattaques par ingénierie sociale.

Pour ce faire, nous avons commencé à construire un nouveau « framework » que l'on nommera « *Neurocyber Framework* » et qui pose le premier cadre théorique et méthodologique de notre démarche fondatrice pour une cybersécurité cognitive.

Ce framework une fois développé dans sa version « progicialisé » pourrait servir de test de référence pour toutes les organisations qui souhaitent évaluer le risque psychologique de leurs collaborateurs en simulation de cyberattaques par ingénierie sociale.

Notre **Neurocyber Framework** est composé de 8 critères d'évaluation qui sont tous des facteurs de vulnérabilité psychologique face aux cyberattaques par ingénierie sociale :



NEUROCYBER FRAMEWORK © (copyright)	Bruno Teboul © (copyright)										
Facteurs de vulnérabilité psychologique face aux cyberattaques		**Scoring**									**Evaluations**
Ouverture : la volonté d'expérimenter de nouvelles choses.		1	2	3	4	5	6	7	8	9	10
Conscience : favorise les normes, fait preuve de maîtrise de soi et l'autodiscipline et la compétence.		1	2	3	4	5	6	7	8	9	10
Extraversion : être plus convivial, extraverti et interactif avec plus de monde.		1	2	3	4	5	6	7	8	9	10
Agréabilité : être coopératif, désireux d'aider les autres et croire à la réciprocité.		1	2	3	4	5	6	7	8	9	10
Neuroticisme : tendance à ressentir des sentiments négatifs, de la culpabilité, le dégoût, la colère, la peur et la tristesse.		1	2	3	4	5	6	7	8	9	10
Auto-efficacité : c'est-à-dire la capacité à gérer des événements inattendus.		1	2	3	4	5	6	7	8	9	10
Niveau de formation : diplômes, études, formations professionnelles.		1	2	3	4	5	6	7	8	9	10
Expertise : niveau de connaissances en « cyber ».		1	2	3	4	5	6	7	8	9	10

- **1. Ouverture : la volonté d'expérimenter de nouvelles choses.**
- **2. Conscience : favorise les normes, fait preuve de maîtrise de soi et l'autodiscipline et la compétence.**
- **3. Extraversion : être plus convivial, extraverti et interactif avec plus de monde.**
- **4. Agréabilité : être coopératif, désireux d'aider les autres et croire à la réciprocité.**
- **5. Neuroticisme : tendance à ressentir des sentiments négatifs, de la culpabilité, le dégoût, la colère, la peur et la tristesse.**
- **6. Auto-efficacité : c'est-à-dire la capacité à gérer des événements inattendus.**
- **7. Formation : niveau de formation (études, formations professionnelles).**
- **8. Expertise : niveau de connaissances en « cyber ».**

Les résultats des premiers tests permettront de dresser une première typologie psychologique de profils plus ou moins à risque face aux cyberattaques. Ces tests pourront être reconduits autant de fois que nécessaire dans le temps, afin de constater une évolution ou non des psycho-types, au sein d'une organisation.

Conclusion

Nous arrivons au terme de notre recherche qui avait pour but d'expliquer l'urgence d'opérer un changement de paradigme en cybersécurité, afin de prendre pleinement le tournant « cognitif » de la cybersécurité. Il s'agissait en effet de démontrer les errements actuels sur le diagnostic des cyberattaques par ingénierie sociale.

Notre incapacité à lutter efficacement contre les cyberattaques (par ingénierie sociale), ouvre la voie à une nouvelle orientation scientifique possible : l'orientation neurocognitive de la cybersécurité.

L'objectif était de bâtir les prolégomènes à toute cybersécurité future comme science (fondée sur l'articulation du raisonnement emprunté à la médecine : diagnostic, thérapeutique et pronostic), en proposant comme cadre théorique et méthodologique les neurosciences cognitives.

Les neurosciences cognitives produisent des méthodes sophistiquées, productrices d'images attrayantes, elles sont de plus en plus affirmées comme explicatives des phénomènes parmi les plus complexes qui soient : les processus d'émergence de la pensée, dans toutes ses composantes.

L'esprit humain est ainsi fait qu'il ne peut se mettre à faire de la science, ou organiser un savoir, sans une vision ou une croyance sur le monde. Rappelons ce que Kuhn appelle un « paradigme », il le définit comme un « *ensemble des croyances, des valeurs reconnues et des techniques qui sont communes aux membres d'une communauté scientifique donnée* », ou bien un « ensemble des exemples ou solutions d'énigmes dans une discipline ».

Or, si l'on peut expérimenter quotidiennement l'intérêt d'une telle matrice disciplinaire (autre nom du paradigme) permettant de poser les termes possibles d'un problème donné et d'y trouver des solutions, notons que celles-ci, cependant, restent locales (c'est-à-dire globalement limitées au paradigme ou à l'une de ses sous-parties) et éventuellement provisoires (c'est-à-dire, comme dirait Popper, falsifiables).

Toutefois, le changement de paradigme en cybersécurité est nécessaire pour construire une approche scientifique, pour lutter efficacement contre la recrudescence des cyberattaques par ingénierie sociale : comme nous l'avons montré, en mobilisant l'arsenal théorique et conceptuel d'une part (neurosciences cognitives) et en développant un cadre expérimental d'autre part, grâce à des tests psychologiques (« neurocyber framework ») qui permettront de comprendre les biais, les heuristiques, les traits psychologiques responsables de nos erreurs, de nos failles cognitives qu'exploitent les cybercriminels quotidiennement.

La neuro-cybersécurité (ou cybersécurité cognitive) donnera naissance à de nouveaux outils, à de nouveaux produits, au bénéfice des acteurs de la sécurité informatique et aux services des organisations publiques et privées, cibles de toutes les attaques « psychologiques » dans le cyberespace. Elle engendrera de nouvelles vocations, de nouveaux débouchés pour cette industrie et créera de nouveaux métiers appliqués à la cybersécurité : « *neuroscientist* »,

« *cognitive scientist* », « *neurocognitive analyst* », dont les compétences et les expertises s'arracheront à prix d'or !

La revanche des sciences humaines et sociales est en marche, l'ère des neuroscientifiques va s'imposée progressivement et sonnera sans doute le glas des promesses déçues de l'intelligence artificielle. La toute-puissance des « data scientists » sera fragilisée par les explorateurs de la conscience humaine et de la psychologie cognitive, qui contribueront à l'avènement d'une société numérique plus résiliente, plus sûre, plus éthique et à visage humain.

Bibliographie

Al'Absi, M., Hugdahl, K., and Lovallo, W. R. (2002). Adrenocortical stress responses and altered working memory performance. Psychophysiology 39, 95–99. doi: 10.1111/1469-8986.3910095

Cho, J.-H., Cam, H., and Oltramari, A. (2016). "Effect of personality traits on trust and risk to phishing vulnerability: modeling and analysis," in 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA) (San Diego, CA: IEEE), 7–13.

Darwish, A., El Zarka, A., and Aloul, F. (2012). "Towards understanding phishing victims' profile," in 2012 International Conference on Computer Systems and Industrial Informatics (Sharjah: IEEE), 1–5. doi: 10.1109/ICCSII.2012.6454454

Digman, J. M. (1997). Higher-order factors of the big five. J. Pers. Soc. Psychol. 73:1246. doi: 10.1037/0022-3514.73.6.1246

Donnellan, M. B., and Robins, R. W. (2009). "The development of personality across the lifespan," in The Cambridge Handbook of Personality Psychology,

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annu. Rev. Psychol. 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629

Gupta, S., Singhal, A., and Kapoor, A. (2016). "A literature survey on social engineering attacks: phishing attack," in 2016 International Conference on Computing, Communication and Automation (ICCCA) (Greater Noida: IEEE), 537–540. doi: 10.1109/CCAA.2016.7813778

Halevi, T., Lewis, J., and Memon, N. (2013). "A pilot study of cyber security and privacy related behavior and personality traits," in Proceedings of the 22nd International Conference on World Wide Web (Singapore: ACM), 737–744. doi: 10.1145/2487788.2488034

Halevi, T., Memon, N., and Nov, O. (2015). Spear-phishing in the wild: a real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. SSRN Electron. J. doi: 10.2139/ssrn.2544742.

Jalali, M. S., Bruckes, M., Westmattelmann, D., and Schewe, G. (2020). Why employees (still) click on phishing links: investigation in hospitals. J. Med. Internet Res. 22:e16775. doi: 10.2196/16775

Jansen, J., and Leukfeldt, R. (2016). Phishing and malware attacks on online banking customers in the Netherlands: a qualitative analysis of factors leading to victimization. Int. J. Cyber Criminol. 10:79. doi: 10.5281/zenodo.58523

Kahneman, D. (2011). Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux.

Kimball, D. R., and Holyoak, K. J. (2000). “Transfer and expertise,” in The Oxford Handbook of Memory, eds E. Tulving, and F. I. M. Craik (New York, NY: Oxford University Press), 109–122.

Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments. J. Pers. Soc. Psychol. 77:1121. doi: 10.1037/0022-3514.77.6.1121

Kumaraguru, P., Acquisti, A., and Cranor, L. F. (2006). “Trust modelling for online transactions: a phishing scenario,” in Proceedings of the 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (Markham, ON: ACM), 11. doi: 10.1145/1501434.1501448

Lupien, S. J., McEwen, B. S., Gunnar, M. R., and Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nat. Rev. Neurosci. 10:434. doi: 10.1038/nrn2639

Mather, M., and Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. Perspect. Psychol. Sci. 6, 114–133. doi: 10.1177/1745691611400234

Redmiles, E. M., Chachra, N., and Waismeyer, B. (2018). “Examining the demand for spam: who clicks?” in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal, ON: ACM), 212. doi: 10.1145/3173574.3173786

Van Schaik, P., Jeske, D., Onibokun, J., Coventry, L., Jansen, J., and Kusev, P. (2017). Risk perceptions of cyber-security and precautionary behaviour. Comput. Hum. Behav. 75, 547–559. doi: 10.1016/j.chb.2017.05.038

Vishwanath, A., Harrison, B., and Ng, Y. J. (2018). Suspicion, cognition, and automaticity model of phishing susceptibility. Commun. Res. 45, 1146–1166. doi: 10.1177/0093650215627483

Wang, J., Herath, T., Chen, R., Vishwanath, A., and Rao, H. R. (2012). Phishing susceptibility: An investigation into the processing of a targeted spear phishing email. IEEE Trans. Profess. Commun. 55, 345–362. doi: 10.1109/TPC.2012.2208392

Wright, R. T., and Marett, K. (2010). The influence of experiential and dispositional factors in phishing: an empirical investigation of the deceived. J. Manage. Inform. Syst. 27, 273–303. doi: 10.2753/MIS0742-12222 70111