



UNIVERSITÉ DE TECHNOLOGIE DE BELFORT-MONTBÉLIARD

Psychologie de la persuasion et conception d'un dispositif d'évaluation

Application des modèles de persuasion (ELM, Dissonance, Nudge) aux comportements de cybersécurité

WALLE Armand
FISE Informatique - MR52

Table des matières

Introduction.....	5
I. De l'influence sociale aux interfaces numériques.....	6
1. Les fondements de la psychologie sociale et de l'influence sociale.....	6
2. Les paradigmes contemporains de la persuasion.....	7
A. Le modèle de probabilité d'élaboration (ELM) et la surcharge cognitive.....	7
B. Ingénierie comportementale : Captologie et Nudges.....	8
C. Le micro-ciblage psychologique.....	8
3. La théorie de l'engagement et la dissonance cognitive.....	8
A. Dissonance cognitive et rationalisation.....	9
B. De la soumission librement consentie à la communication engageante.....	9
C. L'hypocrisie induite comme levier de changement.....	10
II. Analyse critique des dispositifs expérimentaux existants.....	11
1. Évaluation de la vulnérabilité cognitive via les interfaces (Protocole A).....	11
A. Analyse du design : Le piège mathématique.....	12
B. Interprétation des résultats et efficacité du design.....	12
2. Dissonance cognitive et réduction des rationalisations (Protocole B).....	13
A. Analyse de l'impact immédiat (Taille d'effet et Valeur p).....	13
B. Preuve de l'engagement à long terme (Test longitudinal).....	14
3. Le rôle des émotions dans l'engagement (Protocole C).....	14
A. L'effet d'interaction : L'importance du cadrage.....	15
B. Culpabilité ou Honte : L'analyse de médiation.....	16
4. Bilan : Vers un protocole mixte.....	16
III. Conception et Opérationnalisation du protocole de recherche.....	17
1. Genèse et architecture du protocole expérimental mixte.....	17
2. Définition de l'échantillon et ciblage comparatif.....	17
3. Déroulement du scénario et conception du matériel.....	18
A. Le pré-test : Élicitation des préférences.....	18
B. L'intervention psychologique.....	18
C. La tâche distractive (Filler Task).....	18
D. Le piège et l'architecture de choix.....	19
4. Identification et contrôle des variables.....	20
IV. Modélisation des hypothèses et stratégie d'analyse statistique.....	22
1. Construction et justification des hypothèses de recherche.....	22
2. Plan d'analyse et traitement des données.....	23
A. Traitement de la Variable Dépendante 1 (Le choix binaire).....	23
B. Traitement de la Variable Dépendante 2 (Les échelles de Likert).....	23
3. Critères d'interprétation et de validité.....	23
Conclusion.....	25
Bibliographie.....	26
Index des Figures.....	28
Figures indépendantes des sections :	28
Section 3 :	28

Introduction

Chaque jour, nous prenons des milliers de décisions face à nos écrans pour notre travail ou vie personnelle, souvent persuadés de garder le contrôle de nos actions. En réalité, nos environnements numériques s'appuient massivement sur des mécanismes de persuasion invisibles, minutieusement intégrés à leur design.

L'informatique moderne a dépassé la simple création d'outils pour basculer dans l'ingénierie comportementale. Les réseaux sociaux et plateformes en ligne sont maintenant capables d'orienter nos habitudes à notre insu. C'est précisément à l'intersection entre le développement technique et la psychologie humaine que se situe ce travail.

Face à ces manipulations invisibles, notamment en matière de cybersécurité, les modèles de prévention classiques s'avèrent inefficaces, car ils négligent les processus psychologiques profonds qui guident nos comportements.

Pour y répondre, notre cheminement s'articulera autour de quatre axes majeurs. Dans un premier temps, nous construirons le socle théorique en retraçant l'évolution des modèles d'influence sociale et de persuasion, de l'ELM (modèle de probabilité d'élaboration) jusqu'aux théories de l'engagement et de la dissonance cognitive. Ensuite, nous analyserons de manière critique des protocoles expérimentaux récents ayant éprouvé ces concepts, afin d'en extraire les meilleures pratiques méthodologiques. Fort de cette revue, la troisième partie sera consacrée au cœur de notre travail : la conception d'un protocole expérimental mixte et inédit, appliqué au contexte étudiant de l'UTBM, croisant la manipulation d'interfaces (Dark Patterns) et l'induction d'hypocrisie. Enfin, nous modéliserons nos hypothèses de recherche et détaillerons la stratégie d'analyse statistique.

I. De l'influence sociale aux interfaces numériques

1. Les fondements de la psychologie sociale et de l'influence sociale

Pour appréhender les mécanismes de persuasion déployés dans les environnements numériques, il est indispensable de revenir aux fondements théoriques de l'influence et faire un point sur l'état de l'art. Ces concepts s'enracinent dans la psychologie sociale. Selon Jhangiani et Tarry (2022) ainsi que Daffin et Lane (2020), cette discipline se définit comme l'étude scientifique de la manière dont les pensées, les émotions et les comportements des individus sont influencés par la présence réelle, imaginaire ou implicite d'autrui.

Née au croisement de la psychologie, de la sociologie et de la philosophie sociale, cette discipline possède un positionnement épistémologique particulier. Elle se distingue de la sociologie en privilégiant l'analyse des processus psychologiques individuels, tels que la perception, le jugement ou la motivation, plutôt que de se limiter aux seules structures macrosociales. Parallèlement, elle s'écarte de la psychologie de la personnalité en se centrant davantage sur l'effet des situations que sur des traits de caractère stables (Daffin et Lane, 2020 ; Dalhousie University, 2020). L'idée centrale qui structure la psychologie sociale est donc l'approche situationnelle : le comportement d'un individu est toujours co-déterminé par l'interaction entre ses caractéristiques personnelles (sa cognition) et son environnement social (Stangor, 2017).

Au fil de son histoire, de ses premières études expérimentales à la fin du 19e siècle jusqu'à sa consolidation institutionnelle dans les années 1970-1990, la psychologie sociale a structuré ses recherches autour de plusieurs grands ensembles thématiques fondamentaux pour la compréhension de l'influence :

- L'étude des attitudes et de l'influence sociale, qui analyse la formation des opinions, la persuasion, le conformisme et l'obéissance aux normes sociales.
- La cognition sociale, qui décrypte la perception d'autrui et les biais dans le jugement social, une thématique largement documentée par Jhangiani et Tarry (2022).
- L'identité sociale et les relations intergroupes, illustrées par la théorie de l'identité sociale de Tajfel et Turner (1979). Ce modèle démontre que les individus tirent une part importante de leur identité de leurs appartenances à des groupes, ce qui dicte leurs comportements et peut favoriser des biais ou des préjugés (SimplyPsychology, 2023).

Aujourd'hui, face aux mutations technologiques, la psychologie sociale se caractérise par une forte spécialisation et une diversification méthodologique (expériences en

laboratoire, enquêtes en ligne, analyses de bases de données massives). Elle s'oriente désormais massivement vers des problématiques appliquées (Daffin et Lane, 2020). Dans le contexte de notre étude, il est crucial de noter que la psychologie sociale contemporaine analyse de plus en plus les effets des réseaux sociaux et des interfaces numériques sur la conformité aux normes, la diffusion d'informations et les comportements en ligne (Jhangiani et Tarry, 2022). C'est précisément cette transition de l'influence sociale classique "en personne" vers une ingénierie psychosociale appliquée aux interfaces numériques qui nous conduit à explorer les paradigmes modernes de la persuasion.

2. Les paradigmes contemporains de la persuasion

La persuasion contemporaine se distingue radicalement des approches de communication classiques. À l'ère de la quatrième révolution industrielle, l'influence ne s'exerce plus uniquement via des messages de masse (télévision, affichage), mais par le biais de dispositifs interactifs, invisibles et hautement personnalisés. Pour comprendre comment nos attitudes sont modifiées dans ces environnements numériques, la psychologie sociale s'appuie sur de nouveaux paradigmes théoriques et d'ingénierie comportementale.

A. Le modèle de probabilité d'élaboration (ELM) et la surcharge cognitive

Le Modèle de Probabilité d'Élaboration (ELM), initialement théorisé par Petty et Cacioppo, reste le cadre fondamental pour comprendre le traitement de l'information persuasive. Ce modèle postule que la persuasion opère via deux voies distinctes : la voie centrale, qui requiert une réflexion logique et une analyse critique des arguments, et la voie périphérique, activée lorsque les ressources cognitives sont faibles. Dans ce second cas, l'individu se fie à des raccourcis mentaux, appelés heuristiques, pour prendre une décision rapide (SimplyPsychology, 2023).

Aujourd'hui, face à l'infobésité, la pression temporelle et la fatigue décisionnelle inhérentes aux interfaces numériques, c'est la voie périphérique qui domine très largement. Les arguments logiques ont laissé place à des signaux superficiels. Les utilisateurs s'appuient massivement sur des "preuves sociales" (le nombre de likes, de partages ou d'abonnés) ou sur l'esthétique de l'interface (sceaux de sécurité, qualité visuelle) pour valider une information ou un comportement. Cette vulnérabilité cognitive explique pourquoi la conception des interfaces (UI/UX) est devenue une arme de persuasion redoutable.

B. Ingénierie comportementale : Captologie et Nudges

La modification systématique des comportements par les interfaces a donné naissance à la captologie, l'étude des ordinateurs en tant que technologies persuasives. Selon Fogg (2009), l'adoption d'un comportement numérique obéit à une équation précise : elle nécessite la convergence simultanée d'une motivation, d'une capacité à agir (facilité de l'action) et d'un déclencheur (trigger). Les notifications push, le scroll infini ou les récompenses aléatoires sont autant de déclencheurs conçus pour créer des habitudes et maintenir l'engagement de l'utilisateur.

Parallèlement, l'économie comportementale a popularisé le concept de "Nudge" (coup de pouce), massivement transposé dans le domaine numérique. Le Nudge modifie l'architecture des choix pour orienter la décision de l'utilisateur sans jamais interdire d'options (Weinmann, Schneider et Brocke, 2016). Cela se traduit par des cases pré-cochées, des mises en avant par la couleur ou des principes de rareté artificielle. Cependant, lorsque ces mécanismes sont détournés pour manipuler l'utilisateur contre ses propres intérêts (par exemple, rendre un désabonnement excessivement complexe), on bascule dans le domaine des "Dark Patterns" (motifs obscurs), qui feront l'objet de notre protocole expérimental.

C. Le micro-ciblage psychologique

Enfin, la persuasion numérique a franchi un nouveau cap en fusionnant la psychologie sociale et le Big Data. Contrairement à la persuasion traditionnelle, le micro-ciblage psychologique exploite les traces numériques laissées par les utilisateurs (historique de navigation, géolocalisation, interactions sociales) pour déduire des profils de personnalité très précis, souvent basés sur le modèle "Big Five" ou OCEAN.

Comme le démontrent Matz et al. (2017), cette technique permet aux algorithmes d'adapter non seulement le contenu du message, mais également sa forme visuelle et le moment de son affichage, afin de correspondre exactement à la vulnérabilité psychologique de l'utilisateur à l'instant T . Ce paradigme d'hyper-personnalisation clôturera la transition de l'influence sociale. Nous sommes passés d'une persuasion par l'argumentation à une persuasion tacite, intégrée au cœur même du code, du design et du fonctionnement des environnements que nous consultons quotidiennement.

3. La théorie de l'engagement et la dissonance cognitive

Pour comprendre pourquoi les individus maintiennent des comportements en ligne même lorsqu'ils sont contraires à leurs intérêts, il est essentiel de coupler l'étude de la persuasion aux théories de l'engagement et de la dissonance cognitive. Ce couple conceptuel, incontournable en psychologie sociale, démontre que ce ne sont pas

toujours nos idées qui dictent nos actions, mais l'inverse : nos actes modifient à posteriori nos croyances (SimplyPsychology, 2023).

A. Dissonance cognitive et rationalisation

Théorisée initialement par Leon Festinger, la dissonance cognitive désigne l'inconfort psychologique ressenti par un individu lorsque son comportement explicite entre en contradiction avec ses valeurs, croyances ou attitudes personnelles. Cet état de tension provoque un éveil aversif (arousal) que l'individu va naturellement chercher à réduire.

Le paradigme historique de Festinger, dit de la soumission forcée, montre que si une personne est amenée à réaliser un acte contraire à ses attitudes sans y être formellement obligée par une forte récompense ou une menace, elle ressentira une forte dissonance. Pour réduire cette souffrance, l'individu déclenche un mécanisme de "rationalisation cognitive" : il modifie a posteriori son attitude pour l'aligner avec l'acte qu'il vient de commettre. Aujourd'hui, ce mécanisme est par exemple utilisé dans la "gamification" des interfaces, où la réalisation de micro-tâches amène l'utilisateur à s'auto-persuader de l'utilité d'un service pour justifier le temps qu'il y a investi.

B. De la soumission librement consentie à la communication engageante

Dans le sillage de Festinger, les travaux de Kiesler, puis de Joule et Beauvois, ont formalisé la théorie de l'engagement. Celle-ci postule que les individus ne sont pas engagés par leurs idées, mais par leurs actes, à la condition stricte que ces actes soient réalisés dans un contexte perçu de liberté (Courbet, Halimi-Falkowicz et Joule, 2013). Sans cette "illusion de liberté", il n'y a pas d'engagement profond, et donc pas de dissonance en cas de comportement incohérent.

Ce principe a donné naissance à la "communication engageante", massivement utilisée en santé publique et sur le web. Le principe est d'associer un message persuasif à la demande d'un "acte préparatoire" simple et peu coûteux (comme répondre à un quiz ou signer une charte avant d'accéder à un contenu). Le fait d'agir d'abord, librement, augmente de façon très significative l'acceptation du message principal et favorise un changement de comportement pérenne. Des chercheurs ont notamment transposé ces techniques (comme le pied-dans-la-porte) sur Internet, démontrant qu'un petit acte d'engagement numérique (ex: signer une pétition en ligne) augmentait l'acceptation d'une requête plus coûteuse par la suite (Guéguen et Jacob, 2001).

C. L'hypocrisie induite comme levier de changement

L'évolution la plus moderne combinant engagement et dissonance est le paradigme de l'hypocrisie induite (Induced Hypocrisy Paradigm - IHP), développé par Elliot Aronson. Ce protocole d'ingénierie psychologique est conçu pour bloquer la rationalisation et forcer l'alignement des actions sur les valeurs (Stone et Fernandez, 2008).

Ce mécanisme repose sur deux étapes successives et indispensables :

- I. L'engagement normatif : On demande d'abord à la personne de s'engager publiquement en faveur d'une norme sociale (par exemple, rédiger un texte sur l'importance éthique des mots de passe robustes).
- II. La saillance des transgressions : Immédiatement après, on confronte le sujet à ses propres manquements récents face à cette même norme.

La collision de ces deux étapes révèle le statut "d'hypocrite" du sujet, ce qui menace son intégrité morale et génère une dissonance cognitive extrême. La seule issue pour l'individu, incapable de rationaliser face à l'évidence de son propre discours, est de modifier radicalement son comportement futur pour le conformer à la norme qu'il a publiquement défendue. Ce paradigme de l'hypocrisie induite sera la clé de voûte de notre conception expérimentale visant à contrecarrer la manipulation algorithmique par un éveil éthique.

II. Analyse critique des dispositifs expérimentaux existants

Pour répondre à notre problématique consistant à identifier et opérationnaliser les leviers d'engagement et de persuasion, il est crucial d'étudier comment ces concepts théoriques ont été traduits en dispositifs de mesure concrets par la communauté scientifique. Pour cela, nous avons sélectionné trois protocoles expérimentaux récents et complémentaires. Le premier permet de mesurer quantitativement la manipulation par l'interface (la "maladie"). Le second démontre comment la dissonance cognitive peut réduire les justifications néfastes des utilisateurs (le "remède"). Le troisième nuance cette approche en soulignant l'importance cruciale des émotions dans l'engagement. Cette analyse critique aiguillera et permettra la création de notre propre dispositif hybride.

1. Évaluation de la vulnérabilité cognitive via les interfaces (Protocole A)

Pour mesurer l'efficacité de la persuasion périphérique (ELM) dans un environnement numérique, nous nous appuyons sur l'étude de Bogliacino et al., qui ont développé une méthode expérimentale rigoureuse appelée le "Transaction Test". Ce protocole apporte une preuve mathématique de la manipulation algorithmique.

Paramètres	Définition dans le cadre de l'étude
Population ciblée	N = 7430 participants européens (Expérience 1 en ligne) et N = 409 (Expérience 2). Rémunération réelle (points convertibles en euros) pour forcer l'attention.
Étape 1 (Pré-test)	Élicitation des préférences : Évaluation de la valeur monétaire précise (WTA) exigée par le sujet pour accepter la revente de ses données (de 0 à 5 euros).
Variables Indépendantes (VI)	Charge Cognitive : Temps de décision libre (Délai motivé) VS Pression temporelle (perte d'argent par seconde écoulee). Architecture : Neutre VS Information cachée VS Jeu sur les émotions (TWE) VS Personnalisation.

Variable Dépendante (VD)	Incohérence des choix : Le sujet choisit-il l'offre réduisant la protection de ses données pour une remise financière <i>inférieure</i> à la valeur qu'il a lui-même déclarée à l'Étape 1 ?
Résultats majeurs	Les Dark Patterns font passer l'incohérence de base de 37,80 % à 47,25 % (sans pression) et à 50,89 % (sous pression temporelle).

A. Analyse du design : Le piège mathématique

L'innovation de ce protocole réside dans sa Variable Dépendante. L'objectif n'est pas de demander à l'utilisateur s'il se sent manipulé, mais de le prouver mathématiquement. Lors du choix final entre le Package 1 (sécurisé) et le Package 2 (données vendues), la réduction offerte par le Package 2 est fixée par l'algorithme à exactement 2,50 euros *en dessous* de la valeur attribuée à la vie privée par le sujet à l'Étape 1.

Ainsi, si le sujet choisit l'offre réduite, il subit une perte financière mathématique nette. Cela caractérise une violation de l'axiome des préférences révélées, suggérant que l'interface est parvenue à biaiser sa rationalité économique (voie centrale de l'ELM).

B. Interprétation des résultats et efficacité du design

Les résultats de l'Expérience 1 démontrent la puissance de la persuasion périphérique. Même sans pression temporelle, la simple présence de Dark Patterns a fait bondir le taux d'incohérence des choix. Les techniques de l'"Information cachée" (+12,25 points) et de la "Personnalisation" (+9,84 points) se sont révélées particulièrement efficaces pour induire l'utilisateur en erreur. De plus, la saturation de l'attention aggrave cette vulnérabilité : sous pression temporelle, l'incohérence dépasse les 50 %.

Cependant, l'Expérience 2 apporte une nuance fondamentale pour notre recherche. En testant une interface "Bright Pattern" (un Nudge utilisant les mêmes ressorts émotionnels visuels, mais pour valoriser la protection des données), l'incohérence a chuté à 20,5 %. Cela prouve que le design périphérique n'est pas qu'un outil de manipulation néfaste, il peut être inversé pour devenir un levier d'engagement éthique. Cette mesure d'incohérence binaire (Cohérent/Incohérent) servira de socle à notre propre conception.

Il convient toutefois de nuancer ces résultats : l'expérience reposant sur un échantillon en ligne et une rémunération par points convertibles, la pression financière simulée peut différer d'une véritable transaction avec les propres deniers de l'utilisateur.

2. Dissonance cognitive et réduction des rationalisations (Protocole B)

Si le "Transaction Test" démontre l'efficacité de la manipulation, il est indispensable d'explorer les mécanismes psychologiques capables de la contrer. Dans le domaine de la cybersécurité, l'obstacle majeur à la persuasion est l'utilisation par les employés de "techniques de neutralisation" (des rationalisations ou excuses). Le protocole de Siponen, Puhakainen et Vance a été sélectionné car il prouve empiriquement qu'une intervention basée sur la dissonance cognitive peut briser ces barrières mentales.

Paramètres	Définition dans le cadre de l'étude
Population ciblée	N = 80 employés d'une multinationale. Expérience terrain (Field Experiment) quasi-expérimentale.
Temporalité	Trois mesures via questionnaires : Pré-test (avant), Post-test 1 (immédiatement après), Post-test 2 (suivi à 3 semaines).
Variables Indépendantes (VI)	Groupe Contrôle (N = 21) : Formation classique sur les mots de passe. Groupe Expérimental (N = 62) : Formation + Intervention Dissonance (le formateur confronte et démonte les excuses habituelles).
Variables Dépendantes (VD)	1. Intention d'utiliser des mots de passe forts. 2. Intention d'utiliser 8 techniques de neutralisation (ex: "Déni de blessure", "Défense de nécessité"). Mesure sur échelles de Likert.
Résultats majeurs	Baisse significative de 5 des 8 techniques de neutralisation testées et hausse pérenne de l'intention sécuritaire.

A. Analyse de l'impact immédiat (Taille d'effet et Valeur p)

L'intérêt de cette étude réside dans la robustesse de ses analyses statistiques prouvant que la dissonance opère un véritable changement cognitif. Les résultats du Post-test 1 révèlent un impact psychologique majeur de l'intervention.

Pour évaluer la pertinence de ce changement, les auteurs ont utilisé la mesure de la "Taille d'Effet" (Eta-carré, noté η^2). Cette métrique indique quelle proportion du changement de comportement est directement attribuable à l'intervention. L'étude rapporte des η^2 allant de 0.040 à 0.086 pour la réduction des excuses, ce qui représente un effet moyen à large en sciences sociales. Plus remarquable encore, l'augmentation de l'intention d'utiliser des mots de passe sécurisés a généré un η^2 de 0.096 (frôlant les 10 % de variance directement expliquée par l'éveil de la dissonance).

La fiabilité de ces données est confirmée par le seuil de significativité statistique. Avec une "Valeur p" largement inférieure au seuil standard de 5% ($p < 0.01$), la probabilité que ce succès soit dû au hasard est inférieure à 1 %. Par ailleurs, le test ANCOVA ("Score F" de 7.48) confirme que la différence de comportement entre le groupe Dissonance et le groupe Contrôle surpasse largement les simples variations naturelles de personnalité.

B. Preuve de l'engagement à long terme (Test longitudinal)

L'euphorie d'une formation s'estompant généralement vite, les chercheurs ont réalisé un Post-test 2, trois semaines plus tard. L'analyse statistique (Test t pour échantillons indépendants) comparant les scores du Post-test 1 et du Post-test 2 a marqué tous les résultats d'écarts de la mention "N.S." (Non-Significatif).

En science expérimentale, un résultat non-significatif est souvent perçu comme un échec, mais dans un suivi longitudinal, c'est une victoire. Cela prouve mathématiquement qu'il n'y a eu aucune atténuation de l'effet dans le temps. La prise de conscience générée par la dissonance cognitive a bloqué la rationalisation de manière pérenne. Ce protocole nous indique qu'il est indispensable de mesurer ces justifications a posteriori via des échelles de Likert.

Ce protocole présente néanmoins une limite importante liée à son échantillon (N=80 employés d'une seule multinationale). Ce contexte très spécifique pose la question de la généralisation des résultats à un public plus hétérogène ou moins inséré dans une culture d'entreprise stricte.

3. Le rôle des émotions dans l'engagement (Protocole C)

Bien que la dissonance cognitive se révèle efficace (Protocole B), elle n'est pas applicable aveuglément. L'éveil psychologique qu'elle provoque peut générer des réactions opposées selon l'émotion ressentie. Le protocole de Ryoo et Kim, qui étudie l'induction d'hypocrisie dans le cadre d'une campagne RSE (anti-cyberharcèlement), introduit la notion fondamentale de "Focus Régulateur".

Paramètres	Définition dans le cadre de l'étude
Population ciblée	Études en ligne via Amazon MTurk. N = 176 (Étude 1) et N = 293 (Étude 2).
Variables Indépendantes (VI)	Design 2x2 croisé. Hypocrisie : Présente (engagement + rappel des inactions passées) VS Absente. Focus Régulateur : Message "Promotion" (vocabulaire axé sur l'idéal) VS Message "Prévention" (vocabulaire axé sur le risque).
Médiateurs testés	Mesure des émotions ressenties post-dissonance : Culpabilité (Guilt) VS Honte (Shame).
Variables Dépendantes (VD)	Intention de soutenir la campagne (WTS) et intention de faire un don (WTD) mesurées sur 7 points.
Résultats majeurs	Effet d'interaction massif : L'hypocrisie fonctionne très bien avec la "Promotion", mais échoue totalement avec la "Prévention".

A. L'effet d'interaction : L'importance du cadrage

Une interaction survient quand l'effet d'une variable dépend entièrement d'une autre. L'Analyse de Variance (ANOVA) de cette étude révèle une interaction spectaculaire ($F = 18.77, p < 0.001$).

L'induction d'hypocrisie s'est avérée être un amplificateur de persuasion massif (+1.6 point sur la volonté de donner), mais uniquement lorsque l'interface adoptait un ton de "Promotion" (ex: "Bâtissons une cyber-utopie"). L'intervalle de confiance pour ce groupe (95% CI = [0.48, 1.13]) ne contenant pas la valeur zéro, l'effet positif est statistiquement garanti. À l'inverse, l'association de l'hypocrisie avec un bouton d'action formulé sous forme de menace (Focus "Prévention") a totalement neutralisé l'effet de l'intervention (l'intervalle de confiance traversant le zéro).

B. Culpabilité ou Honte : L'analyse de médiation

Pour expliquer ce blocage, les auteurs ont utilisé l'outil statistique PROCESS (Modèle 8 de Hayes) pour tracer la "chaîne de causalité" émotionnelle. Ils ont prouvé que la dissonance cognitive (l'Hypocrisie) n'entraîne pas directement l'action. Elle génère d'abord une émotion, qui devient la courroie de transmission vers le comportement.

L'étude démontre que faire ressentir de la *Culpabilité* (une focalisation sur l'acte) motive l'individu à corriger sa faute (médiateur significatif, effet d'Approche). Cependant, combiner l'hypocrisie à un message anxiogène génère énormément de *Honte* (une focalisation sur l'identité). Cette surcharge émotionnelle, destructrice pour l'ego de l'utilisateur, agit comme un frein à main (médiateur non significatif). La seule défense psychologique de l'individu est alors la fuite (Évitement) en fermant la page.

Ces échantillons de population invitent cependant à la prudence. En effet, la plateforme Amazon MTurk est souvent utilisée par des "répondants professionnels". Leur niveau d'attention ou leur sensibilité préalable aux enjeux éthiques abordés pourraient introduire un biais. Bien que les résultats de cette étude soient probants, ces éléments limitent leur généralisation.

4. Bilan : Vers un protocole mixte

L'analyse de ces trois protocoles dicte l'architecture de notre propre dispositif expérimental. Le Transaction Test (Protocol A) prouve que l'architecture de choix dicte les comportements sous charge cognitive, mais il n'explore pas les défenses psychologiques. Le protocole de Neutralisation (Protocol B) prouve que la dissonance déconstruit les fausses excuses de manière pérenne, mais il n'évalue pas cette résilience face à une interface délibérément manipulatoire. Enfin, l'étude sur le Focus Régulateur (Protocol C) nous met en garde sur la nécessité de cadrer l'induction d'hypocrisie pour générer de la culpabilité réparatrice, et non de la honte.

Pour répondre à notre problématique, nous devons concevoir un protocole hybride. En tant que chercheur découvrant ce domaine, cette démarche expérimentale vise à synthétiser les tenants et aboutissants des protocoles précédents pour proposer une approche adaptée. Il s'agira ainsi de soumettre les participants au piège du Transaction Test (Dark Patterns), tout en évaluant si une induction préalable d'hypocrisie (dissonance) peut agir comme un "bouclier mental" suffisant pour bloquer la soumission à la voie périphérique.

L'intégration de ces limites méthodologiques, et tout particulièrement le manque de validité écologique de certains échantillons en ligne, justifie notre choix de cibler une population étudiante bien définie (UTBM) au travers d'une application fictive ancrée dans ses habitudes locales.

III. Conception et Opérationnalisation du protocole de recherche

Dans le prolongement direct des limites et synergies identifiées précédemment, il s'agit de confronter dans cette partie, la force manipulatoire du design (l'attaque) à la résilience psychologique de l'utilisateur (la défense). Notre démarche méthodologique consiste à ancrer ce "choc" dans un environnement numérique contrôlé, mesurable et écologiquement valide pour notre population cible.

1. Genèse et architecture du protocole expérimental mixte

Pour tester l'efficacité de la communication engageante face à la manipulation algorithmique, notre dispositif repose sur la fusion conceptuelle de deux méthodologies distinctes.

D'une part, nous reprenons l'architecture du *Transaction Test* modélisé par Bogliacino et al. (2023). Cette méthode permet de piéger mathématiquement l'utilisateur : plutôt que de lui demander s'il se sent manipulé, on observe s'il valide une transaction financièrement perdante par rapport à ses propres exigences initiales. D'autre part, nous intégrons le traitement par dissonance cognitive éprouvé par Siponen, Puhakainen et Vance (2020). L'enjeu est d'utiliser l'hypocrisie induite pour briser les mécanismes de rationalisation avant même que l'utilisateur ne soit confronté au choix final.

Afin de garantir l'implication des participants (la validité écologique), le scénario doit s'ancrer dans leur quotidien. L'expérience simule donc la création d'un profil sur une application fictive nommée "Bons Plans UTBM", proposant des réductions étudiantes (Fnac, Uber Eats, etc.). L'utilisateur devra choisir entre un profil sécurisé mais contraignant, et un profil express, rémunérateur mais intrusif pour ses données personnelles.

2. Définition de l'échantillon et ciblage comparatif

Le protocole cible spécifiquement les étudiants de l'Université de Technologie de Belfort-Montbéliard (UTBM). L'utilisation d'un logiciel de création d'enquêtes conditionnelles (type Qualtrics ou Limesurvey) permettra d'automatiser l'assignation aléatoire des participants.

La valeur ajoutée de cet échantillon réside dans la comparaison de deux groupes naturels distincts, permettant de vérifier si le bagage technique influence la susceptibilité aux Dark Patterns :

- Filière Informatique : Ce panel représente les "experts techniques". Il s'agit d'observer si leurs compétences en cybersécurité les protègent intuitivement de la manipulation visuelle, ou si, à l'inverse, ils rationalisent leur prise de risque par un excès de confiance (par exemple en affirmant utiliser sciemment un "mot de passe poubelle").
- Filière Ergonomie, Design et Ingénierie Mécanique : Ce panel représente les "experts en usages". Sensibilisés à la conception d'interfaces (UI/UX), ils pourraient décrypter le Dark Pattern plus facilement, ou au contraire, justifier une faille de sécurité par le besoin de fluidité de navigation et le refus d'une friction ergonomique.

3. Déroulement du scénario et conception du matériel

Le protocole suit un séquençage strict en quatre étapes, pensé pour masquer les véritables intentions de l'étude jusqu'à la dernière seconde.

A. Le pré-test : Élicitation des préférences

Le contexte de l'application "Bons Plans UTBM" est posé. Le participant doit répondre à une question d'évaluation financière précise : *"Quelle valeur MINIMUM en cartes cadeaux par mois exigeriez-vous pour accepter de créer votre compte en mode Express (sécurité faible et partage de données) ?"*. Il choisit un palier (2, 5, 10, 15, 20 euros) ou refuse catégoriquement. Cette valeur est enregistrée secrètement par le système.

B. L'intervention psychologique

Tous les sujets visionnent une vidéo officielle de prévention (Cybermalveillance.gouv.fr) sur les mots de passe. Ensuite, l'échantillon est scindé. Le groupe de contrôle passe à la suite. Le groupe expérimental subit le traitement par dissonance : il doit répondre par Oui ou Non à un questionnaire d'hypocrisie confrontant (ex: *"Avez-vous déjà utilisé un mot de passe très simple par flemme de le retenir ?"*). L'objectif est de générer la culpabilité réparatrice identifiée dans la littérature.

C. La tâche distractive (Filler Task)

Pour éviter que le participant ne garde consciemment en tête la somme choisie, une tâche de diversion est imposée. Il doit remplir un rapide formulaire socio-démographique (âge, genre, semestre, filière). Cela vide la mémoire de travail immédiate.

D. Le piège et l'architecture de choix

L'étudiant doit finaliser la création de son compte. On lui propose le "Profil Sécurisé" (aucun gain) ou le "Profil Express" (avec un bonus immédiat de 5 euros). Si l'étudiant avait exigé 15 euros au début de l'étude et qu'il clique sur le profil à 5 euros, le mécanisme opère : l'interface a vraisemblablement influencé son choix à son insu. Selon l'assignation, cet écran est présenté soit via une "Interface Neutre" (transparente et équitable), soit via une "Interface Dark Pattern" combinant le *Trick Wording* (jeu sur l'urgence) et l'interférence visuelle (bouton vert proéminent, risques écrits en gris clair).



Figure 1 : Interface incluant des Dark Patterns (Source : Réalisation personnelle)



Figure 2 : Interface neutre (Source : Réalisation personnelle)

4. Identification et contrôle des variables

L'architecture s'appuie sur un plan d'expérience factoriel 2x2. Les composantes manipulées et observées sont définies ainsi :

Variables Indépendantes (VI) - Manipulées et contrôlées :

- VI 1 : Filière d'étude (Informatique vs EDIM).
- VI 2 : Traitement Psychologique préalable (Sensibilisation neutre vs Induction d'hypocrisie).
- VI 3 : Architecture de Choix finale (Interface Neutre vs Interface Dark Pattern).

Variables Dépendantes (VD) - Mesurées :

- VD 1 : L'incohérence du choix. Mesure binaire (0 ou 1) issue du Transaction Test. Détermine si l'étudiant a cédé à l'interface en acceptant une offre financièrement inférieure à ses propres critères de sécurité.
- VD 2 : L'intensité des techniques de neutralisation. Mesure quantitative post-expérience (via des échelles de Likert de 1 à 5). Elle évalue la propension du sujet à justifier son comportement par de fausses excuses (dénier de blessure, défense de nécessité, etc.).

La matrice ci-dessous illustre le croisement de nos variables générant les quatre conditions de test. La variable de filière (VI 1) s'applique de manière transversale à l'ensemble des groupes pour garantir la représentativité.

Paramètres expérimentaux	Groupe 1 (Contrôle)	Groupe 2 (Expérimental A)	Groupe 3 (Expérimental B)	Groupe 4 (Choc des effets)
VI 1 : Population assignée	Info & EDIM (aléatoire)	Info & EDIM (aléatoire)	Info & EDIM (aléatoire)	Info & EDIM (aléatoire)
VI 2 : Phase de Préparation (Traitement)	Contrôle (Vidéo de sensibilisation)	Contrôle (Vidéo de sensibilisation)	Dissonance (Vidéo + Hypocrisie)	Dissonance (Vidéo + Hypocrisie)
VI 3 : Phase de Décision (Interface)	Interface Neutre (transparente)	Dark Pattern (TWE + Interférence)	Interface Neutre (transparente)	Dark Pattern (TWE + Interférence)
Variables Dépendantes (Mesures)	VD 1 : Incohérence (0/1) VD 2 : Scores de Likert	VD 1 : Incohérence (0/1) VD 2 : Scores de Likert	VD 1 : Incohérence (0/1) VD 2 : Scores de Likert	VD 1 : Incohérence (0/1) VD 2 : Scores de Likert

IV. Modélisation des hypothèses et stratégie d'analyse statistique

Concevoir un protocole expérimental rigoureux n'est qu'une étape. Pour que ce dispositif ait une véritable valeur scientifique, il doit s'accompagner d'attentes claires et d'un plan d'analyse capable de transformer les comportements observés en données chiffrées incontestables. L'objectif de cette quatrième partie n'est pas de deviner les résultats, mais de formuler des hypothèses déduites de notre revue de littérature et de définir précisément les outils statistiques qui permettront de les valider ou de les rejeter lors de la future passation.

1. Construction et justification des hypothèses de recherche

Nos hypothèses se divisent en trois axes : l'effet de l'interface, l'effet de la psychologie, et les spécificités de notre population cible.

- Hypothèse 1 (H1) : L'impact du bagage technique sur la rationalisation. Nous postulons que face au Dark Pattern, les étudiants en Informatique et en EDIM pourraient présenter des taux d'échec (incohérence) similaires, mais qu'ils justifieront leur chute de manière radicalement différente. Forts de leur bagage technique, les étudiants en Informatique rationaliseront davantage leur choix par des arguments de contournement (ex : affirmer utiliser un "mot de passe poubelle" pour limiter le risque). À l'inverse, les étudiants en EDIM (orientés design) justifieront leur comportement inadapté par la défense de nécessité liée à la praticité (le refus d'une friction ergonomique excessive lors de la création d'un compte).
- Hypothèse 2 (H2) : L'effet manipulateur direct de l'interface. En nous basant sur le *Transaction Test* (Bogliacino et al., 2023), nous postulons que l'interface Dark Pattern (qui combine le *Trick Wording* émotionnel et l'interférence visuelle) augmentera significativement le taux de choix du Profil Express (choix incohérent) par rapport à l'interface neutre, soutenant l'hypothèse d'une forte efficacité de la persuasion périphérique sous charge cognitive.
- Hypothèse 3 (H3) : La neutralisation des excuses par la dissonance. Dans la lignée des travaux de Siponen, Puhakainen et Vance (2020), nous postulons que les groupes ayant subi le traitement par hypocrisie induite (Groupes 3 et 4) présenteront des scores d'utilisation des techniques de neutralisation (VD 2 mesurée en post-test) significativement plus faibles que les groupes de contrôle, prouvant la déconstruction des fausses excuses.
- Hypothèse 4 (H4) : L'effet "bouclier mental" (L'interaction). C'est le cœur de notre étude. Face à l'interface trompeuse, l'induction préalable de dissonance cognitive agira comme un mécanisme de défense. Le Groupe 4 (qui subit l'attaque du Dark Pattern après avoir été éveillé éthiquement) présentera un

taux d'incohérence significativement inférieur à celui du Groupe 2 (qui subit le Dark Pattern seul). L'éveil de l'éthique personnelle devrait suffire à bloquer la soumission aux indices périphériques.

2. Plan d'analyse et traitement des données

Notre protocole génère deux types de variables très différentes qui nécessitent des traitements statistiques spécifiques.

A. Traitement de la Variable Dépendante 1 (Le choix binaire)

L'incohérence du choix issu du *Transaction Test* produit une donnée binaire (0 = cohérent, 1 = incohérent). On ne peut pas calculer de moyenne sur ce type de donnée. Pour valider H2 et H4, nous devons utiliser un Test du Khi-Deux (χ^2). Ce test permettra de vérifier si la différence de proportion d'étudiants piégés entre l'interface neutre et le Dark Pattern est mathématiquement réelle ou si elle est simplement due au hasard. Pour aller plus loin et isoler le "poids" exact du Dark Pattern par rapport à l'âge ou au sexe, une modélisation par régression logistique sera pertinente.

B. Traitement de la Variable Dépendante 2 (Les échelles de Likert)

Les scores mesurant l'intensité des techniques de neutralisation (les excuses) sont des données quantitatives (de 1 à 5). Pour valider H1 et H3, nous procéderons à des comparaisons de moyennes. Si nous comparons simplement le groupe "Vidéo neutre" au groupe "Dissonance", un Test T de Student pour échantillons indépendants semble convenable. En revanche, pour croiser l'effet de la filière (Info vs EDIM) avec le type de traitement psychologique de manière simultanée, nous utiliserons une Analyse de Variance à deux facteurs (ANOVA 2-way).

3. Critères d'interprétation et de validité

Enfin, il est crucial de définir à l'avance les critères de succès de notre dispositif.

- **Le seuil de significativité (la valeur p) :** En sciences sociales, nous considérerons une hypothèse comme validée uniquement si la valeur p (probabilité que le résultat soit dû au hasard) est strictement inférieure à 0,05.

- **L'effet d'interaction** : Comme l'ont démontré Ryoo et Kim (2024), l'effet croisé de plusieurs variables est l'indicateur le plus riche en psychologie sociale. L'ANOVA nous permettra de vérifier si l'impact de la dissonance dépend entièrement de l'interface présentée, c'est-à-dire de confirmer mathématiquement le succès de notre Hypothèse 4.
- **La taille d'effet et les intervalles de confiance** : Dire qu'un résultat est "significatif" ne veut pas dire qu'il est "massif" dans la vie réelle. Nous calculerons donc l'Eta-carré (η^2) pour les ANOVA et nous analyserons les intervalles de confiance à 95% (95% CI). Si ces intervalles ne croisent pas le zéro, cela nous garantira la fiabilité de nos futurs résultats sur l'efficacité de la dissonance cognitive comme remède au design manipulateur.

Conclusion

L'informatique moderne a dépassé le stade d'outil de productivité. Face à des environnements numériques conçus pour exploiter nos failles cognitives via la voie périphérique, les modèles de prévention classiques, qui s'adressent uniquement à notre raison, s'avèrent inefficaces.

Ce mémoire s'est attaché à déconstruire ces dynamiques en croisant le design persuasif (Dark Patterns) et la psychologie de l'engagement (dissonance cognitive). En nous appuyant sur les limites des protocoles existants, nous avons conçu un outil expérimental hybride, spécifiquement adapté au contexte étudiant de l'UTBM. Ce dispositif permet de mesurer l'impact frontal entre une manipulation visuelle agressive (le piège du *Transaction Test*) et une défense psychologique interne (l'induction d'une hypocrisie réparatrice).

Ce travail s'arrête au seuil de l'expérimentation, avec une méthodologie prête à être déployée. L'enjeu de la passation future sera de confronter notre modèle théorique à la réalité du terrain, avec des résultats qui pourront s'avérer surprenants :

- Scénario validant : Les données prouvent que l'induction d'hypocrisie agit comme un "bouclier mental" efficace contre les Dark Patterns, et confirment que le bagage technique (Info vs EDIM) génère des techniques de rationalisation distinctes.
- Scénario invalidant : Les données révèlent que l'architecture de choix manipulateur est si puissante qu'elle court-circuite l'éveil éthique, précipitant même les experts techniques dans le piège.

Qu'ils valident ou invalident nos hypothèses, ces futurs résultats seront cruciaux. Ils permettront de mesurer la véritable résilience de l'utilisateur face au design persuasif. Comprendre cette mécanique est la condition préalable pour espérer, à terme, retourner ces outils à notre avantage et concevoir une cybersécurité préventive fondée sur la psychologie humaine plutôt que sur la seule contrainte technique.

Bien que ce protocole ait été construit avec rigueur, il convient d'en souligner les limites inhérentes. Tout d'abord, notre échantillon étant restreint aux étudiants de l'UTBM (filiales Informatique et EDIM), les résultats obtenus ne pourront être totalement généralisés à l'ensemble de la population ou à d'autres tranches d'âge. De plus, notre dispositif s'appuie sur une situation expérimentale simulée (la création d'un profil sur une application fictive) et sur des mesures déclaratives (échelles de Likert pour les justifications). Il subsiste donc un risque de biais de désirabilité sociale, les participants pouvant rationaliser différemment dans un contexte réel impliquant leurs véritables données personnelles.

Bibliographie

LIVRES EN LIGNE

DAFFIN, Lee W. Jr., LANE, Carrie. *Principles of Social Psychology* [en ligne]. 2nd ed. Washington State University, 2020, 238 p. Disponible sur : <https://opentext.wsu.edu/social-psychology/> (consulté le 18 mars 2026)

JHANGIANI, Rajiv S., TARRY, Hammond. *Principles of Social Psychology* [en ligne]. 1st International H5P Edition. BCcampus, 2022, 736 p. Disponible sur : <https://opentextbc.ca/socialpsychology/> (consulté le 21 mars 2026)

STANGOR, Charles. *Principles of Social Psychology* [en ligne]. 1st International Edition. eCampusOntario, 2017, 620 p. Disponible sur : <https://openlibrary-repo.ecampusontario.ca/jspui/handle/123456789/527> (consulté le 19 mars 2026)

CHAPITRES DE LIVRE EN LIGNE

DALHOUSIE UNIVERSITY. *What is Social Psychology?* In: *Introduction to Psychology & Neuroscience* [en ligne]. Dalhousie University Libraries, 2020. Disponible sur : <https://digitaleditions.library.dal.ca/intropsychneuro/chapter/what-is-social-psychology/> (consulté le 24 mars 2026)

COURBET, Didier, FOURQUET-COURBET, Marie-Pierre, BERNARD, Françoise et al. *Communication persuasive et communication engageante pour la santé : Favoriser des comportements sains avec les médias, Internet et les serious games*. In: BLANC, Nathalie (dir.). *Publicité et Santé : des liaisons dangereuses ? Le point de vue de la psychologie* [en ligne]. In Press, 2013, p. 15. Disponible sur : https://archivesic.ccsd.cnrs.fr/sic_01076750v1 (consulté le 23 mars 2026)

ARTICLES EN LIGNE

MATZ, Sandra C., KOSINSKI, Michal, NAVE, Gideon et al. *Psychological targeting as an effective approach to digital mass persuasion*. In: *Proceedings of the National Academy of Sciences* [en ligne]. 28 novembre 2017, vol. 114, n° 48, p. 12714-12719. Disponible sur : <https://www.pnas.org/doi/10.1073/pnas.1710966114> (consulté le 22 mars 2026)

WEINMANN, Markus, SCHNEIDER, Christoph, BROCKE, Jan vom. *Digital Nudging*. In: *Business & Information Systems Engineering* [en ligne]. 2016, vol. 58, p. 433-436. Disponible sur : <https://link.springer.com/article/10.1007/s12599-016-0453-1> (consulté le 26 mars 2026)

GUÉGUEN, Nicolas, JACOB, Céline. *Fund-Raising on the Web: The Effect of an Electronic Foot-in-the-Door on Donation*. In: *CyberPsychology & Behavior* [en ligne]. 2001, vol. 4, n° 6, p. 705-709. Disponible sur : https://www.researchgate.net/publication/11559126_Fund-Raising_on_the_Web_The_Effect_of_an_Electronic_Foot-in-the-Door_on_Donation (consulté le 21 mars 2026)

STONE, Jeff, FERNANDEZ, Nicholas C. *To Practice What We Preach: The Use of Hypocrisy and Cognitive Dissonance to Motivate Behavior Change*. In: *Social and Personality Psychology Compass* [en ligne]. 13 mars 2008, vol. 2, n° 2, p. 1024-1051. Disponible sur : <https://compass.onlinelibrary.wiley.com/doi/10.1111/j.1751-9004.2008.00088.x> (consulté le 25 mars 2026)

BOGLIACINO, Francesco, PEJSACHOWICZ, Leonardo, LIVA, Giovanni et al. *Testing for Manipulation: Experimental Evidence on Dark Patterns*. In: *SocArXiv* [en ligne]. 2023. Disponible sur : https://doi.org/10.31235/osf.io/sqt3j_v1 (consulté le 9 avril 2026)

SIPONEN, Mikko, PUHAKAINEN, Petri, VANCE, Anthony. *Can individuals' neutralization techniques be overcome? A field experiment on password policy*. In: *Computers & Security* [en ligne]. 2020, vol. 88, p.101617. Disponible sur : https://jyx.jyu.fi/jyx/Record/jyx_123456789_65920 (consulté le 12 avril 2026)

RYOO, Yuhosua, KIM, WooJin. *Approach versus avoidance: A self-regulatory perspective on hypocrisy induction in anti-cyberbullying CSR campaigns*. In: *Journal of Business Ethics* [en ligne]. 2024, vol. 189, n° 2, p. 345-364. Disponible sur : https://www.researchgate.net/publication/368926901_Approach_versus_Avoidance_A_Self-Regulatory_Perspective_on_Hypocrisy_Induction_in_Anti-Cyberbullying_CSR_Campaigns (consulté le 18 avril 2026)

PAGES SUR SITE INTERNET

SIMPLYPSYCHOLOGY. *Social Identity Theory in Psychology*. In: *SimplyPsychology* [en ligne]. 2023. Disponible sur : <https://www.simplypsychology.org/social-identity-theory.html> (consulté le 15 mars 2026)

CONFÉRENCES / CONGRÈS / COLLOQUES ET ACTES EN LIGNE

FOGG, B. J. *A Behavior Model for Persuasive Design* [en ligne]. Proceedings of the 4th International Conference on Persuasive Technology, Association for Computing Machinery. New York, 2009. Disponible sur : <https://dl.acm.org/doi/10.1145/1541948.1541999> (consulté le 16 mars 2026)

Index des Figures

Figures indépendantes des sections :

- Page de garde : DANILYUK, Pavel. *Sans titre*. In: Pexels [en ligne]. Disponible sur : <https://images.pexels.com/photos/7654411/pexels-photo-7654411.jpeg> (Consulté le 4 juin 2026).

Section 3 :

- Figure 1 : Interface incluant des Dark Patterns (Source : Réalisation personnelle) p. 19
- Figure 2 : Interface neutre (Source : Réalisation personnelle) p. 20