# Tester ou ne pas tester?

Une analyse critique des lacunes des tests A/B



# E ARCHITECTURE DE L'INFORMATION

MEMOIRE
PROFESSIONNEL DE
MASTER II EN SCIENCES
HUMAINES ET SOCIALES,
MENTION
ARCHITECTURE DE
L'INFORMATION

SOUS LA DIRECTION DE M. AURELIEN TABARD

ANNEE UNIVERSITAIRE 2014 / 2015













# **RESUME**

L'expression « test A/B » renvoie à une méthode d'optimisation de la conversion clairement identifiée. A l'opposé, le terme « tests A/B » désigne un ensemble de démarches. Les tests A/B permettent de mettre en concurrence plusieurs variations d'un design afin d'identifier celle qui s'avère la plus performante. Les différentes versions sont testées sur des utilisateurs d'un dispositif en situation d'usage réel, sans qu'ils en soient informés. Dans le même temps, leurs actions sont enregistrées sous forme de données numériques quantitatives. Les solutions les plus avancées ont rendu accessible la création de tests A/B au plus grand nombre. De ce fait, des profils d'utilisateurs variés sont amenés à recourir à ce type de solution logicielle, notamment des professionnels de l'expérience utilisateur, du design, de l'architecture de l'information et du webmarketing. Le marché des tests A/B est extrêmement concurrentiel et en forte croissance. L'augmentation du taux de pénétration des différentes solutions s'explique par les retours sur investissements considérables que permettent d'atteindre les tests A/B. Dans un contexte économique morose, ils présentent l'avantage de permettre de faire progresser sensiblement ses performances, tout en maintenant ses investissements à un niveau constant. En résumé, la valeur des tests A/B réside dans le fait qu':

- ils s'adressent à tous les profils et secteurs d'activité.
- ils sont simples d'usage et peu coûteux.
- ils peuvent faire progresser significativement la performance et quantifier précisément les variations enregistrées.
- ils permettent de tester des hypothèses sur des utilisateurs authentiques dans des contextes d'usages réels.

Grâce à leurs nombreuses qualités, les tests A/B ont acquis un statut d'outil incontournable, notamment chez les professionnels du webmarketing. Pourtant, ils ne font pas l'unanimité. En effet, il existe des voix dissonantes dans le concert de louanges dont les tests A/B font l'objet. Ces opinions divergentes constituent une invitation à une analyse critique et constructive à laquelle est convié le lecteur. Les exposés successifs des points forts des tests A/B, de leurs lacunes, des modes opératoires des professionnels pour les contourner, le tout augmenté de la glose de l'auteur, seront l'occasion pour le lecteur de s'interroger sur le bien-fondé d'un recours aux tests A/B.

# TABLE DES MATIERES

Résumé	2
Table des matières	. 3
Introduction	4
Un concept à la mode qui renvoie à de multiples outils et usages	5
Des promesses séduisantes	5
Des motifs d'utilisation variés	7
Une offre d'outils pléthorique	8
Une utilisation qui ne va pas de soi	10
L'architecte de l'information et les tests A/B sur le Web	11
S'interroger sur les limites des tests A/B pour savoir quand les utiliser 1	12
Deux grandes faiblesses à examiner	12
Des outils perfectibles	14
De nombreuses limitations techniques	14
Des indicateurs de performance peu fiables	14
Des palliatifs nombreux mais insatisfaisants	16
Contrôler les populations pour mieux tester ?	19
Des données quantitatives pour principale source d'informations	21
De nombreuses lacunes à surmonter	21
Le pari hasardeux de la multiplication des sources de données	24
Entrer en contact pour mieux tester ?	29
Une démarche en question 3	33
Un concept éthiquement critiquable	33
Des risques de dérives importants	33
La transparence : une pratique marginale	36
Respecter les directives de la CNIL pour mieux tester ?	37
Une démarche qui requiert du temps	<del>1</del> 0
Le temps, une source majeure d'erreurs d'interprétation	<del>1</del> 0
Les calculateurs d'échantillon : une réponse unique mais insatisfaisante4	13
Allonger les durées pour mieux tester ?	45
Conclusion	<b>17</b>
Bibliographie5	50

## INTRODUCTION

Le monde ne s'est toujours pas remis de la crise économique de 2007. L'environnement économique demeure incertain, en proie à des changements massifs causés notamment par la poursuite de la bascule vers le numérique. Pour se prémunir de ces effets, les acteurs du système économique cherchent davantage à comprendre les motivations et attentes de leurs clients. De ce fait, il existe une forte demande pour des outils et méthodologies basées sur des retours utilisateurs. Au XXIème siècle, opter pour un pilotage par la donnée <sup>1</sup> est perçu comme un facteur de réduction des risques lors de l'élaboration et/ou de l'amélioration d'un produit ou d'un service. Or, grâce aux progrès continus de l'informatique depuis la seconde moitié du XXème siècle, l'acquisition et le stockage des données sont devenus accessibles au plus grand nombre. La base d'utilisateurs potentiels et les opportunités d'exploitation s'en sont trouvées décuplées <sup>2</sup>. La convergence de ces trois tendances :

- 1. Le besoin de minimiser les risques qui s'explique par le contexte économique morose.
- 2. La croyance selon laquelle la prise en compte du besoin utilisateur est le moyen le plus efficace de parvenir à cet objectif.
- 3. La possibilité de capter, de stocker et de traiter de façon automatisée des données sous forme numériques à un faible coût.

a abouti à l'instauration d'un pilotage des actions et décisions reposant sur des outils informatiques qui exploitent des données.

Cette tendance, s'applique à l'ensemble des activités humaines contemporaines. Elle est apparue dans le secteur des services en ligne <sup>3</sup> qui ont vu le jour grâce aux nouvelles technologies de l'information et des communications. En effet, dans l'écosystème numérique la viabilité d'un service est conditionnée à l'instauration d'une relation entre ses concepteurs et ses utilisateurs. Dans ce secteur d'activité, la compréhension des usages et besoins types des utilisateurs repose sur l'exploitation des données générées lors de l'utilisation d'un programme. A tel point, qu'une nouvelle « science des données » <sup>4</sup> a vu le jour. Les professionnels de cette discipline travaillent sur des volumes de données considérables générées par des milliers d'utilisateurs. C'est ainsi qu'est apparu le terme *big data* <sup>5</sup> qui désigne tout à la fois :

- Le volume considérable de données générées et stockées par les institutions publiques, les organismes non gouvernementaux et les entreprises.
- La production extrêmement rapide de données qui découlent du recours à un service par un utilisateur.
- La possibilité de pouvoir exploiter les ensembles de données ainsi générés <sup>6</sup>.

<sup>&</sup>lt;sup>1</sup> Les Anglo-Saxon emploient l'expression « Data-driven process and decisions ».

<sup>&</sup>lt;sup>2</sup> InfoDesign, Putting people first Philips Design: From data to meaning for people,

 $<sup>\</sup>underline{http://www.experientia.com/blog/philips-design-from-data-to-meaning-for-people/}, 4 \ novembre \ 2012.$ 

<sup>&</sup>lt;sup>3</sup> Accessibles grâce à l'Internet, notamment sur le Web.

<sup>&</sup>lt;sup>4</sup> Lin Jimmy, Wongsuphasawat Krist, *Using Visualizations to Monitor Changes and Harvest Insights from a Global-Scale Logging Infrastructure at Twitter*,

http://www.umiacs.umd.edu/~jimmylin/publications/Wongsuphasawat\_Lin\_VAST2014.pdf, 13 novembre 2014.

<sup>&</sup>lt;sup>5</sup> Littéralement les « mégadonnées ». Le terme désigne à la fois la constitution, le stockage et l'exploitation d'ensemble de données si massifs qu'ils en deviennent ingérables avec des outils classiques de gestion des bases de données et de l'information.

<sup>&</sup>lt;sup>6</sup> ESPOSTI SARA DEGLI, « When big data meets dataveillance : the hidden side of analytics », in : *Surveillance & Society*, vol. 12, n° 2, mai 2014, pp. 209-225.

C'est dans ce contexte, qu'au tournant du siècle dernier, une tendance a commencé à se dégager dans le milieu des services en ligne : le *testing* <sup>7</sup>. Puisqu'il était devenu possible de collecter à bas coût des données et de les analyser pour réduire l'incertitude et les risques, le recours aux tests se généralisa. A l'opposé de la morosité de l'économie mondiale, le marché des tests en ligne connait une croissance exponentielle, véritable décalque de celle des big data.

## Un concept à la mode qui renvoie à de multiples outils et usages

Le recours à une démarche de tests peut s'expliquer par des problématiques diverses. Les tests A/B constituent l'une des nombreuses méthodes de tests disponibles sur le Web. Leur spécificité découle de l'existence d'outils et de méthodologies qui leurs sont dédiés. Afin d'apporter une réponse aux multiples cas d'usages existant, une pléthore d'outils est commercialisée sur le marché florissant des tests A/B. Ce dynamisme s'explique par les gains substantiels que le recours aux tests A/B doit permettre d'atteindre grâce à l'analyse de données collectées lors du test.

### Des promesses séduisantes

Pour comprendre l'intérêt et la valeur ajoutée des tests A/B, il est nécessaire d'en donner une définition. Bien que l'expression soit employée couramment dans les domaines du *Webmarketing* <sup>8</sup> et du *Web Analytique* <sup>9</sup> il n'en existe pas d'acceptation commune du fait de la diversité des usages qui y sont associés. Voici les principales caractéristiques des tests A/B :

- Les tests A/B sont avant tout une démarche générique, qui peut varier fortement en fonction des méthodologies employées.
- L'usage des tests A/B répond au souci d'identifier le produit ou le service fournissant les meilleurs résultats.
- Les tests A/B peuvent consister à mettre en concurrence plusieurs variations d'un même produit/service ou plusieurs produits/services concurrents.
- Un test A/B est une expérimentation au cours de laquelle un unique paramètre est modifié d'une variation à une autre. Par opposition un test multivarié a pour objet l'analyse des effets de la combinaison de plusieurs paramètres au sein d'une même variation.

Dans le langage courant ; l'emploi de l'expression test[s] A/B désigne tout à la fois la démarche de test dans son ensemble, mais aussi un type de test en particulier. En outre, elle renvoie à la fois aux tests A/B, multivarié, mais aussi à ceux visant à mettre à l'épreuve des solutions ou des designs concurrents. Le champ d'application des tests A/B est extrêmement vaste. Cependant, il comporte une limite : les contenus testés doivent être délivrés sous forme numérique car il s'avère indispensable d'être en mesure de

<sup>8</sup> Secteur d'activité dédié à l'amélioration de la performance d'une organisation via des actions d'optimisation sur ses différents dispositifs connectés au Web.

 $<sup>^{7}</sup>$  La traduction littérale de ce nom anglais signifie « essais », « tests ».

<sup>&</sup>lt;sup>9</sup> Activité qui consiste à analyser l'audience et les comportements des visiteurs d'un site Web.

quantifier automatiquement l'impact des modifications déployées. A l'heure actuelle, les tests A/B sont principalement utilisés sur les sites Web, bien que les tests d'applications mobiles et de prototypes occupent une place sans cesse croissante. L'engouement pour les tests A/B reste pour l'heure cantonné aux professionnels de l'Internet, en particulier ceux du Web. En effet, l'évaluation de l'impact d'un test effectué sur des canaux de diffusion de l'information plus traditionnels tels que la télévision et la radio s'avère pour l'heure techniquement complexe.

Si les tests A/B ont fait l'objet de nombreuses recherches académiques <sup>10</sup>, la plupart des écrits qui leurs sont consacrés appartiennent à la littérature du marketing Web 11. Ceci s'explique par une caractéristique commune au marketing et aux tests A/B : la recherche de la meilleure performance possible. Or, l'expérience démontre que la modification d'éléments mineurs tels que le libellé, l'emplacement, la taille, ou encore la couleur d'un élément suffisent à générer des résultats conséquents. C'est ainsi qu'en février 2014, Google annonçait avoir augmenté ses revenus de près de 200 millions de dollars en ayant modifié la teinte de bleu des liens publicitaires inclus dans les pages de résultats de son moteur de recherche <sup>12</sup>. La vocation des tests A/B est de déterminer la meilleure solution à un problème donné, sur la base d'éléments clairement définis par l'étude des traces d'utilisation d'un dispositif <sup>13</sup>. La collecte passive de données constitue la principale valeur ajoutée des tests A/B sur des méthodes concurrentes. Les approches traditionnelles reposaient sur des entretiens utilisateurs particulièrement onéreux et dont la conduite ne pouvait être confiée qu'à des spécialistes. En outre, le faible nombre de personnes interrogées et la nature *qualitative* <sup>14</sup> des données recueillies sont suspectés de nuire à l'objectivité et à la fiabilité des analyses qui en découlent. A l'opposé, un unique test A/B peut être effectué sur des milliers d'individus. Seules sont prises en compte des données quantitatives dont la collecte et l'interprétation est confiée à des machines et des algorithmes.

Le crédit dont jouissent les tests A/B tient également au fait que les données sont collectées dans un contexte d'utilisation réel, à l'inverse des expériences et entretiens traditionnels <sup>15</sup> où des participants accomplissent des tâches qui leurs sont dictées en ayant pleinement conscience d'être observés dans un environnement contrôlé <sup>16</sup>. Mener à bien un test A/B ne nécessite pas de compétences techniques particulières puisqu'un programme se charge de l'affichage des modifications et du calcul des résultats. En revanche, réaliser une série d'entretiens utilisateurs ou mettre en concurrence deux solutions par des moyens traditionnels sont des tâches qui nécessitent de l'expérience et des compétences solides. Autre argument de poids en faveur des tests A/B : un test n'est jamais mené en pure perte. En l'absence de test préalable, le choix d'un design ou d'une solution peut s'avérer désastreux. Les tests A/B permettent de s'assurer au minimum qu'il n'y ait pas de régression dans les performances. En cas d'hésitation entre deux choix possibles, ils permettent d'identifier le plus performant et d'éliminer tout risque d'erreur.

<sup>&</sup>lt;sup>10</sup> Pour obtenir le détail des références, voir la section bibliographique de l'article de Scott Young (CF Infra).

<sup>&</sup>lt;sup>11</sup> Abrévitation de World Wide Web en Anglais, littéralement la toile d'araignée mondiale. Il s'agit d'un système hypertexte public fonctionnant sur l'internet.

<sup>&</sup>lt;sup>12</sup> BARANIUK CHRIS, *Internet: How websites are experimenting on you*, <a href="http://www.bbc.com/future/story/20140801-hidden-tricks-that-make-you-click">http://www.bbc.com/future/story/20140801-hidden-tricks-that-make-you-click</a>, août 2014.

<sup>&</sup>lt;sup>13</sup> NIELSEN JAKOB, *Putting A/B Testing in Its Place*, <a href="http://www.nngroup.com/articles/putting-ab-testing-in-its-place/">http://www.nngroup.com/articles/putting-ab-testing-in-its-place/</a>, août 2005.

<sup>14</sup> Données descriptives textuelles. Elles s'opposent aux données quantitatives qui sont des données descriptives numériques.

<sup>&</sup>lt;sup>15</sup> NIELSEN JAKOB, *Ibidem*.

<sup>&</sup>lt;sup>16</sup> Bernheim Brush A.J., *Ubiquitous Computing Fundamentals*, s.l., Chapman and Hall/CRC, 2009, 328 p.

Accessibles et économiques, les tests A/B sont perçus comme une démarche impartiale permettant de définir mathématiquement la meilleure solution possible dans le cadre d'une expérimentation. Ils permettent au minimum de ne pas régresser et promettent d'aider à identifier les axes d'amélioration possibles. Ces qualités expliquent que des professionnels d'horizons divers s'intéressent aux tests A/B.

#### Des motifs d'utilisation variés

Le terme test A/B est employé principalement pour désigner des expérimentations affectant l'ergonomie d'une page ou d'un écran. Cependant, l'expression peut également désigner la mise en concurrence de plusieurs logiciels. Bien que cet usage demeure marginal, il peut s'avérer extrêmement utile. Ainsi, le propriétaire d'un site Web ou d'une application, désireux de tirer profit de son portefeuille de visiteurs, peut décider d'y inclure des espaces publicitaires. Le marché de l'affiliation <sup>17</sup> compte de nombreux acteurs, comment déterminer celui pouvant s'avérer le plus rentable ? En souscrivant à une version d'essai de plusieurs solutions d'affiliation, il devient possible de déterminer celle à même de générer le plus de bénéfices. Dans ce contexte, le motif d'utilisation des tests A/B est la recherche du bénéfice maximum.

Le recours aux tests A/B dans le cadre de l'amélioration continue d'un service et/ou d'un design donnant satisfaction est motivé avant tout par la recherche d'une progression substantielle au prix d'un investissement minime. Exemple : une entreprise désireuse d'augmenter le *taux de conversion* <sup>18</sup> de son formulaire de contact destiné à collecter les coordonnées de prospects décide de lancer un test A/B. La suppression d'une photographie lui a permis d'augmenter de 24% le nombre de prises de contact enregistrées. <sup>19</sup>

Les tests A/B peuvent également être utilisés pour redresser une situation jugée critique. Lors de la campagne présidentielle de 2008 aux Etats-Unis, l'équipe du candidat Barack Obama savait qu'il ne pourrait collecter un volume de fonds comparable à celui du candidat républicain John McCain auprès des entreprises et des particuliers fortunés. Pour égaler les républicains, ils décidèrent de miser sur le nombre et non le montant en cherchant à obtenir un maximum de dons de la part de citoyens ordinaires. L'un des vecteurs d'acquisition ou de transformation <sup>20</sup> de donateurs potentiels était le site Internet du candidat. Cependant, le ratio entre le nombre de visiteurs et celui des personnes ayant déposé une demande de contact était jugé insatisfaisant. La mise en place d'un test multivarié devait permettre d'améliorer significativement les performances initiales du site et de récolter près de 60 millions de dollars supplémentaires <sup>21</sup>.

Le recours aux tests A/B peut également se justifier dans le cadre d'une refonte globale. Dès lors l'objectif n'est plus de tester l'impact d'une modification mineure, mais plutôt des designs antagonistes. L'enjeu est alors double :

- S'assurer que le nouveau design ne présente pas des performances inférieures à celui qu'il est destiné à remplacer.
- Valider que la progression de performance enregistrée est à la hauteur des investissements et des attentes.

<sup>&</sup>lt;sup>17</sup> Technique de Webmarketing autorisant un marchand à diffuser son catalogue de produits ou de publicités sur le site Web d'un tiers : l'affilié.

<sup>&</sup>lt;sup>18</sup> Pourcentage des utilisateurs sur le total réalisant une action donnée.

<sup>&</sup>lt;sup>19</sup> https://whichtestwon.com/case-study/hubspots-lead-generation-landing-page-test/.

<sup>&</sup>lt;sup>20</sup> Le terme a ici une signification similaire à celle de la conversion.

<sup>&</sup>lt;sup>21</sup> http://blog.optimizely.com/2010/11/29/how-obama-raised-60-million-by-running-a-simple-experiment/.

L'utilisation des tests A/B pour la mise en concurrence de designs antagonistes ne fait pas l'unanimité pour plusieurs raisons :

- L'élaboration d'un nouveau design est un processus long, couteux et risqué, à l'opposé des modifications opérées lors de la recherche d'une optimisation.
- Le nombre d'éléments modifiés est tel qu'il s'avère impossible en cas d'échec ou de succès d'en déterminer la cause.

Reste que la recherche d'une rupture dans le design est le seul moyen pour dépasser le phénomène dit du *local maximum* <sup>22</sup> auquel aboutit la mise en place d'une série d'optimisations mineures. C'est ainsi, qu'un auteur put obtenir une progression de 131% du taux d'inscription à sa lettre d'information en testant un nouveau design <sup>23</sup>.

L'utilisation des tests A/B peut également s'expliquer par l'existence d'un conflit au sein d'une organisation. L'évolution d'un design, l'ajout ou le retrait d'une fonctionnalité peuvent être la source de conflits et de tensions. Les tests A/B offrent la possibilité de trancher pour une solution plutôt qu'une autre sur la base d'indicateurs quantitatifs. Les conflits potentiels se trouvent ainsi désamorcés, chacun ayant l'opportunité de comparer les performances de sa propre proposition à celle de ses pairs. La décision finale n'étant pas prise par un être humain, chacune des parties prenantes sait que le risque de partialité et de favoritisme s'en trouve considérablement minoré. A l'occasion d'une prestation d'optimisation du site d'un courtier en assurance, une divergence d'opinion est apparue entre le client et les équipes d'*Empirik* <sup>24</sup>. Le client n'était pas convaincu par la proposition d'ajouter d'un outil de calcul des intérêts comme élément de réassurance sur le site. La mise en place d'un test A/B a été présentée comme une méthode d'évaluation impartiale de la recommandation. Le conflit a ainsi pu être désamorcé.

Enfin, les outils de tests A/B peuvent être utilisés à des fins de personnalisation de l'expérience utilisateur. L'idée est simple : plus l'expérience de navigation d'un visiteur est personnalisée, plus le ratio entre le nombre de visites et de conversion est élevé. Les outils de tests A/B sont en mesure de délivrer un contenu spécifique pour différentes typologies de visiteurs. En fonction de son matériel, de son canal d'acquisition et du nombre de visites à son actif, pour ne citer que les critères les plus utilisés, le visiteur est assigné à une catégorie d'utilisateur et se voit proposer un contenu personnalisé.

Ce bref passage en revue des principaux cas d'usage des tests A/B met en évidence leur subordination à des démarches et problématiques qui les englobent et les dépassent. Il s'agit principalement d'évaluer un design et/ou une architecture de l'infirmation et assez marginalement des tests d'outils et de solution. De ce fait, les principaux utilisateurs des outils de tests A/B sont des professionnels du marketing, du design et des analystes. La complexité des tests mis en place dépend autant de la compétence de l'équipe en charge des tests que des possibilités offertes par l'outil utilisé.

## Une offre d'outils pléthorique

Il n'existe pas de logiciels dédiés aux tests de solutions et d'outils dans la mesure où la distribution aléatoire d'un code simple peut être effectuée sans avoir à recourir à un outil dédié. Avant tout, c'est la possibilité de pouvoir analyser l'impact du déploiement de chacun des outils directement au sein de leur interface respective qui rend inutile le

<sup>&</sup>lt;sup>22</sup> Point à partir duquel l'amélioration d'un design devient impossible sans opérer des modifications d'envergure.

<sup>23</sup> http://blog.hubspot.com/blog/tabid/6307/bid/20569/Why-Marketers-A-B-Testing-Shouldn-t-Be-Limited-to-Small-Changes.aspx

<sup>&</sup>lt;sup>24</sup> Agence Webmarketing basée à Lyon qui emploie l'auteur.

recours à un logiciel de tests A/B. Sur un site Web équipé d'un gestionnaire de balises <sup>25</sup>,

le déploiement du test peut s'effectuer sans même solliciter avoir à un développeur. En revanche, le test d'un design ou d'une architecture l'information requiert obligatoirement d'utiliser un logiciel dédié aux tests A/B. La valeur ajoutée du logiciel réside avant tout dans sa capacité l'impact quantifier des modifications déployées. Les outils de tests A/B ont

Enterprise Rank	Q4 2014	Q3 2014
1	Adobe Target	Adobe Target
2	Monetate	Monetate
3	Qubit	Maxymiser
4	Maxymiser	Optimizely
5	Optimizely	Qubit
6	Sitespect	Sitespect
7	Visual Website Optimizer	Visual Website Optimizer
8	HP Optimost	Webtrends Optimize
9	Webtrends Optimize	HP Optimost
10	Dynamic Yield	AB Tasty

Fig. 1 : Solutions de tests A/B pour lesquelles le nombre d'avis déposé sur la plateforme d'avis Trust Radius est le plus important

Source : Top 10 Fastest Growing A/B Testing Software Products Q4 2014

<a href="https://www.trustradius.com/articles/Top-10-Fastest-Growing-ABTesting-Software-Q414">https://www.trustradius.com/articles/Top-10-Fastest-Growing-ABTesting-Software-Q414</a>.

recours à trois grandes méthodes afin de déployer les modifications apportées à la version d'origine d'un écran ou d'une page :

- La modification côté serveur sur le Web / directement dans le code source de l'application sur mobiles. Il s'agit de l'approche la plus marginale, mais aussi la plus fiable. Elle tend à disparaître du fait de la nécessité d'avoir des compétences en développement afin de mettre en place le code du test A/B. Concrètement, le code modifiant la page / l'écran s'exécute avant l'envoi de la ressource au client, ce qui garantit le contenu de la ressource envoyée. Le logiciel de tests A/B gère l'assignation aléatoire d'un utilisateur à une variation ou à la version d'origine ainsi que le suivi des performances de la variation.
- Sur le Web, il est possible d'utiliser la *redirection http* <sup>26</sup> dans le cas où les modifications apportées ne concernent qu'un volume de page restreint, ou encore si le design à tester est disponible en pré-production. Une version alternative de la page Web d'origine est développée et hébergée sur un serveur Web. Lorsque des visiteurs cherchent à accéder à la ressource d'origine, l'outil de test A/B en redirige une portion vers la version de test de la page. L'outil de test gère la redirection et effectue le suivi des actions utilisateurs sur la version d'origine et alternative de la page.
- La modification côté client sur le Web et sur application. Il n'existe pas de version alternative de la page/de l'écran. La page ou l'écran sont modifiés au moment de de leur chargement par l'outil de tests A/B qui gère également l'assignation aléatoire des utilisateurs à une variation ainsi que le suivi de leurs actions. Cette méthode est la plus répandue, mais aussi la moins efficace d'un point de vue technique. Sur le Web, le rendu des pages peut varier sensiblement d'un navigateur à un autre avec pour conséquence dans les cas les plus extrêmes de rendre impossible l'exécution d'un test. Le principal inconvénient demeure le

<sup>26</sup> La redirection permet de diriger une partie des visiteurs d'une page A vers une page B sans nécessiter d'intervention de la part du visiteur.

9

-

<sup>&</sup>lt;sup>25</sup> En anglais Tag Management. Un code Javascript unique déposé sur l'ensemble des pages d'un site Web permet le déploiement d'un nombre illimité de balises HTML contenant du CSS ou du Javascript. Les nouveaux codes sont ajoutés au moyen d'une interface en ligne. De nouvelles solutions logicielles et fonctionnalités peuvent ainsi être ajoutées à un site par des utilisateurs sans compétences techniques particulières. Le recours aux gestionnaires de balises permet aux équipes marketing de gagner en agilité en évitant de solliciter les départements informatiques.

risque de scintillement. Ce phénomène se produit en cas de modifications nombreuses, ou si l'un des éléments à modifier est situé en bas de la page/de l'écran. Du point de vue de l'utilisateur, le scintillement se manifeste par l'affichage pendant quelques secondes du design d'origine avant que la version modifiée ne prenne sa place. L'attention de l'internaute étant attirée plus spécifiquement sur les éléments modifiés, il existe un risque de fausser les résultats du test.

Malgré ces défauts, l'exécution des modifications côté client reste la méthode de référence utilisée par l'écrasante majorité des outils. Elle a constitué un pré-requis indispensable à l'élargissement de la base des utilisateurs des solutions de tests A/B audelà des seuls profils techniques. Or, les profils les plus susceptibles d'avoir recours aux tests A/B n'ont dans la plupart des cas qu'un bagage technique limité ou inexistant. Déporter l'exécution des modifications côté client permet de proposer une interface *SAAS* incluant un éditeur *WYSIWYG* <sup>28</sup> à partir de laquelle s'effectuent la configuration, le déploiement et l'analyse des résultats. Dès lors, les pré-requis à l'exécution d'un test se limitent à l'inclusion d'un code identique sur l'ensemble des pages/écran d'un(e) site/application sur lequel/laquelle le test devra s'effectuer.

Si le marché des tests A/B est dominé par des solutions spécialisées, en France notamment A/B tasty; Optimizely et Kameleoon <sup>29</sup>, la tendance lourde du marché est à l'inclusion progressive de fonctionnalités de tests A/B directement au sein d'outils de Web analytique dédiés à l'E-mailing (Hubspot), les campagnes publicitaires (Adwords) et l'analyse comportementale (Google Analytics). Outre la personnalisation des contenus qu'elles sont les seules à proposer, les solutions dédiées aux tests A/B se distinguent par leur polyvalence. Surtout, elles proposent des fonctionnalités avancées qui font défaut aux solutions de Web analytique. Il faut citer notamment l'intégration avec des outils tiers pour y injecter les résultats des tests, des algorithmes de calcul des résultats plus avancés, la compatibilité avec les applications mobiles.

La disponibilité d'un nombre considérable de solutions dédiées et l'inclusion progressive de fonctionnalités de tests A/B dans des outils de Web Analytique sont les signes d'un dynamisme incontestable du marché des tests A/B. Considérant le nombre conséquent de démarches au sein desquelles ces tests peuvent trouver leur place et les bénéfices incontestables associés à leur usage, leur utilisation semble aller de soi. Etant données leurs qualités, l'absence d'une utilisation systématique a de quoi étonner.

## Une utilisation qui ne va pas de soi

Une analyse des tests A/B ne saurait être complète sans s'intéresser à leurs lacunes. Dans cette optique, l'examen des défauts des tests A/B doit permettre de définir les cas de figure dans lesquels leur utilisation peut être jugée pertinente et ceux où elle doit être écartée. A cet égard, l'architecte de l'information apparaît comme l'utilisateur type le plus à même d'exploiter à son plein potentiel un outil dédié aux tests A/B. Quant au Web, il constitue le terrain d'expérimentation dominant et le plus complexe à gérer.

-

<sup>&</sup>lt;sup>27</sup> Littéralement Software As A Service. Application dont l'usage ne nécessite pas d'installation sur un poste de travail. La solution logicielle est hébergée sur des serveurs distants accessibles via une connexion Web.

<sup>&</sup>lt;sup>28</sup> Littéralement What I See Is What I Get. Il s'agit d'un éditeur interactif permettant de modifier le code HTML, CSS et Javascript d'une page sans avoir besoin de compétences en développement.

<sup>&</sup>lt;sup>29</sup> http://www.journaldunet.com/solutions/analytics/web-analytics-tag-management-a-b-testing-les-parts-de-marche-en-france/part-de-marche-des-outils-de-tag-management-et-d-a-b-testing.shtml.

Examiner les difficultés à surmonter dans le cadre d'une combinaison de cet utilisateur et de ce périmètre revient à passer en revue les cas d'usages les plus complexes.

#### L'architecte de l'information et les tests A/B sur le Web

Le profil d'un architecte de l'information est intéressant à plusieurs titres. Bien qu'il ne dispose pas de compétences techniques comparables à celles d'un développeur, sa sensibilité aux technologies Web lui permet d'adapter des fonctions et codes existants dans l'optique de les réutiliser pour des tests A/B. Dès lors, les possibilités de paramétrage qui lui sont offertes dépassent de loin celles de la plupart des utilisateurs, notamment des profils purement marketing. En outre, ses connaissances dans les domaines de l'ergonomie et de l'expérience utilisateur lui permettent d'élaborer des tests de façon réfléchie et non pas empirique. En effet, la base d'utilisateurs des solutions de tests A/B est constituée en majorité de spécialistes du Web Marketing. Certes, ces utilisateurs ont recours à l'analyse de données quantitatives lorsqu'ils déterminent les modifications à apporter à une page. Seulement, la pertinence de leurs recommandations reste avant tout dépendante de leur niveau d'expérience. Pour la plupart, ils n'ont jamais eu recours aux données qualitatives qu'ils jugent moins fiables que les données quantitatives. Leur principal défaut réside dans l'objectif qu'ils cherchent à atteindre au travers des tests. En effet, si l'amélioration d'un taux de conversion constitue également un indicateur de réussite pour l'architecte de l'information, elle ne constitue pas pour lui une fin en soi. Ainsi, l'amélioration de l'expérience utilisateur demeure son principal objectif, bien qu'elle s'avère presque impossible à évaluer sur la base d'indicateurs purement quantitatifs. L'architecte de l'information peut être amené, tous comme les professionnels du Webmarketing, à effectuer des tests en vue de sélectionner un outil ou une solution. Il apparaît que l'architecte de l'information, du fait de son champ de compétence élargi et ses multiples domaines de compétences peut utiliser l'ensemble des démarches et fonctionnalités rattachées aux tests A/B. Cependant, la profession d'architecte de l'information demeurant peu répandue à l'heure actuelle en Europe, notamment en France, ce profil ne peut être qualifié d'utilisateur type des tests A/B. Aux Etats-Unis et au Canada où la profession est la mieux implantée, les architectes de l'information du fait de leur statut d'expert constituent une communauté d'experts largement inférieure en nombre aux analystes, experts du marketing ou du design. De ce fait, les exemples et les cas d'usages présentés renverront principalement aux professionnels du marketing, de l'ergonomie et aux analystes et non pas à des individus disposant de connaissances fortes dans plusieurs domaines. Bien qu'ils ne soient pas directement nommés, ce sont les situations auxquelles pourraient être confronté un architecte de l'information sur lesquelles se concentre ce mémoire

Le Web constitue le champ d'expérimentation le plus complexe à appréhender pour toute personne désireuse de réaliser des tests A/B. La problématique de la compatibilité multinavigateurs des modifications déployées au moyen des outils de tests A/B est un enjeu fort mais un objectif difficile à atteindre. Certains facteurs peuvent conduire à complexifier davantage encore le problème, notamment la nature des visiteurs du site testé. Ainsi, dans le cas d'un site destiné à des professionnels, la part de navigateurs obsolètes, notamment Internet Explorer dans ses versions 6 à 9 s'avèrera largement supérieure à la normale. Il peut même arriver que les utilisateurs de ces types de navigateurs représentent la majorité des personnes testées. A l'heure actuelle, seule la solution de l'entreprise Optimizely, le leader mondial sur le marché des outils dédiés aux tests A/B, propose une fonctionnalité de test multi-navigateurs. La prévisualisation des

résultats n'est offerte qu'aux souscripteurs de l'offre la plus onéreuse. En outre, le module ne vérifie pas à proprement parler les écarts de rendus entre navigateurs, il permet simplement d'en avoir une restitution sous la forme d'une image. Outre l'impossibilité de vérifier la bonne exécution d'une modification appliquée de façon dynamique, le nombre considérable de combinaisons possibles rend toute lecture exhaustive des exports impossible.

Le recours à la librairie Javascript jQuery pour effectuer les modifications côté client, à l'exception notable de la solution française Kameleoon, permet de niveler seulement partiellement les écarts entre navigateurs. Cependant, la principale difficulté demeure de vérifier que les modifications s'effectuent correctement sur l'ensemble des pages incluses dans le test. En effet, les scénarios d'usages potentiels sont beaucoup plus nombreux sur le Web que sur une application. Bien qu'il ne s'agisse que d'un aperçu des difficultés à surmonter, il permet de se faire une idée de l'ampleur des problèmes auxquels un architecte de l'information peut se trouver confronté.

## S'interroger sur les limites des tests A/B pour savoir quand les utiliser

L'identification, l'évaluation et la prise en compte des différentes difficultés à surmonter dans le cadre de la réalisation d'un test A/B apparaissent indispensables pour décider s'il convient d'y recourir. En effet, la charge de travail associée à la création d'un test A/B est proportionnelle à sa complexité. La préparation d'un test complexe peut amener une équipe entière à travailler à sa création. Or, plus le temps investi en préparation s'avère conséquent, plus les attentes sur les résultats des tests s'avèrent importantes. Le risque d'échec, pour cause de progression insuffisante s'en trouve démultiplié. Bien qu'indésirable, l'échec d'un test mené par une entreprise pour son propre compte peut être relativisé dans la mesure où les équipes impliquées dans la préparation sont montées en compétences suite à son élaboration. En outre, l'absence de résultats probants permet d'écarter une hypothèse et ce faisant d'augmenter les chances de réussite des tests suivants.

La situation est radicalement différente pour une agence Web ou une société de services en ingénierie informatique. Certes, en cas d'échec leurs salariés ont pu accroitre leur expérience et réduire la liste des hypothèses à tester. Mais au sein de ces structures, l'échec s'accompagne d'une absence de rentabilité puisque l'obligation de résultats impose de poursuivre les tests jusqu'à obtenir des performances jugées satisfaisantes. Le constat est valable pour l'ensemble des tests qui visent à améliorer l'ergonomie d'un dispositif Web, y compris les plus simples. En effet, le nombre de vérifications nécessaires conduit à un volume d'heures de préparation incompressible très élevé.

Afin de préserver la rentabilité d'un projet où il est fait usage des tests A/B et obtenir un retour sur investissement maximal il faut passer en revue le fonctionnement des outils de et des démarches de tests A/B en se demandant : Quelles sont les limites des tests A/B ? Quelles conséquences ces limites ont-elles sur les résultats des tests ? Existe-t-il des solutions viables permettant de les dépasser ?

## Deux grandes faiblesses à examiner

Les tests A/B souffrent de deux grands types de problèmes :

• Des limitations techniques :

- Elles découlent des méthodes de collecte et de structuration de l'information des outils de tests A/B. Les éditeurs de solutions ont une responsabilité non négligeable dans cet état de fait, dans la mesure où ils prennent soin d'éviter d'évoquer les manques de leurs solutions <sup>30</sup>. Les professionnels du secteur des tests A/B ont bien envisagé de recourir aux données stockées dans le CRM de leur entreprise/client pour contourner le problème, mais cette solution n'est pas pleinement satisfaisante. Il apparaît que le meilleur moyen de contourner le problème est encore de contrôler au mieux les caractéristiques des visiteurs exposés aux tests A/B.
- O Non seulement la fiabilité des indicateurs est douteuse, mais elle est encore aggravée par la nature exclusivement quantitative des données collectées. De ce fait, il s'avère presque impossible de déterminer la cause de l'échec ou de la réussite d'un test A/B. Pour contourner le problème, la marche à suivre communément admise repose sur la collecte et l'analyse de données qualitatives en complément des données quantitatives avant et pendant la réalisation du test. La démarche résiste difficilement à l'épreuve des faits, dans la mesure où il s'avère malaisé de relier un répondant à une personne testée. Sans compter que l'analyse de ces nouveaux ensembles de données s'avère ardue et chronophage. Contacter et interroger directement quelques individus est une approche qui semble offrir davantage de souplesse pour des résultats identiques voir supérieurs.
- Des lacunes qui concernent la démarche proprement dite :
  - O La frontière entre les tests A/B et la manipulation émotionnelle s'avère extrêmement ténue. La nécessité de garder secrète la réalisation du test pour ne pas risquer de modifier le comportement des personnes testées peut conduire à de nombreuses dérives. La transparence et l'information des visiteurs et potentiels sujets de test apparaît comme une possible solution mais demeure un phénomène marginal. L'application des directives de la *CNIL* relatives au respect de la vie privée aurait le mérite de normaliser les pratiques entre les sites et d'avoir l'assurance que les visiteurs d'un site acceptent véritablement d'être testés.
  - Les impératifs de rentabilité conduisent à tirer des conclusions trop hâtives sur les performances d'un test. Des facteurs externes pouvant conduire à minorer les résultats sont systématiquement rejetés par les professionnels du secteur des tests A/B. Les outils de calcul disponibles pour évaluer la durée nécessaire d'un test ont pour unité de mesure l'utilisateur et non le temps. L'obtention de résultats pérennes semble n'être envisageable qu'au prix d'un allongement significatif de la durée des tests.

S'être intéressé au profil des utilisateurs des solutions de tests A/B a permis de lister les principaux contextes dans lesquels le recours à un test A/B s'avérait utile. Le fonctionnement technique des grandes familles d'outils de tests A/B a également été passé en revue et des différences significatives ont pu être mises en évidence. La valeur ajoutée des tests A/B a été exposée sans opter pour une approche critique tentant compte de leurs lacunes. Il s'agit à présent de s'y intéresser.

-

<sup>&</sup>lt;sup>30</sup> http://blog.crazyegg.com/2014/06/25/best-testing-software/.

## DES OUTILS PERFECTIBLES

La connaissance des modalités de collecte et de traitement de l'information par un outil s'avère indispensable à la bonne analyse des données collectées. Faute de connaître les grands principes de fonctionnement d'une solution, l'analyste s'expose à des erreurs d'interprétation. Ainsi, si l'unité de mesure de la performance des outils de tests A/B est l'utilisateur, le terme désigne en réalité un navigateur et non pas un individu. De même, l'apparente impartialité des outils de test A/B, garantie par des traitements algorithmiques, peine à masquer les manques induits par l'usage d'indicateurs exclusivement quantitatifs.

## De nombreuses limitations techniques

La plupart des problèmes qui affectent les logiciels de tests A/B ne leurs sont pas spécifiques. Ils concernent l'ensemble des applications Web qui sont affectées par les limites et le fonctionnement des processus de stockage et de transmission d'informations sur le Web. Il est important de les lister pour en tenir compte lors des analyses mais surtout en amont lorsque la décision est prise de recourir à des tests A/B. Connaître les biais inhérents à la collecte des données, c'est se prémunir contre leurs effets les plus néfastes.

## Des indicateurs de performance peu fiables

Le problème principal affectant les solutions de tests A/B disponibles sur le Web est la notion d'utilisateur. En effet, tout visiteur d'un site Web y accède au moyen d'un navigateur. Lorsqu'un visiteur charge pour la première fois une page faisant l'objet d'un test A/B il se voit proposer une version de la page testée, l'original ou une variation, à condition qu'il satisfasse aux critères de ciblage. En effet, lors du paramétrage du test, il est possible de définir deux types de règles :

- Des règles relatives au périmètre, autrement dit au contexte dans lequel le test doit être déclenché. Les principaux critères de ciblage sont l'URL, la présence d'un élément dans le *DOM* <sup>31</sup> de la page ou d'une variable Javascript globale, à moins qu'un événement dynamique ne soit le signal déclencheur.
- Des règles relatives à l'audience, basées sur les caractéristiques des visiteurs du site. Les informations disponibles sont en quasi-exclusivité celles exploitables grâce à l'analyse des en-têtes HTTP:
  - o Provenance
  - Version du navigateur
  - o Adresse IP
  - o Type d'appareil
  - o Résolution d'écran
  - Système d'exploitation
  - o Langue de l'interface
  - o Plugins installés
  - o Nouveau/ancien visiteur

<sup>31</sup> Littéralement Document Object Model. Il s'agit d'une interface permettant à des programmes rédigés dans des langages multiples de lire et/ou de manipuler l'arbre d'un document XML ou HTML.

Cette dernière information, l'outil la déduit de la présence ou de l'absence d'une série de cookies 32 déposée par la solution de tests A/B lors du chargement d'une page Web où elle se trouve incluse. Le nombre, la forme et la nature exacte des informations stockées peuvent être amenés à varier d'un outil à un autre. Cependant, la répartition de ces informations peut être résumée grossièrement de la façon suivante :

- Un cookie contient l'identifiant unique du visiteur.
- Un autre contient les informations relatives à sa session actuelle, notamment la provenance, le nombre de pages consultées et le nombre de venues sur le site.
- Un troisième cookie permet de garder en mémoire pour un test donné l'assignation d'un visiteur à une variation, à la version d'origine ou encore son exclusion.

Le stockage de ces informations dans des cookies les rend peu pérennes. En cas de navigation privée ou d'effacement de l'historique, elles ne sont pas conservées. Si le test a lieu sur deux noms de domaines distincts, elles ne peuvent être partagées du fait des restrictions multi-domaines inhérentes au protocole HTTP, ou alors au prix d'une configuration complexe. Pour peu qu'un même utilisateur en vienne à utiliser différents appareils pour se connecter à un même site, ou simplement plusieurs navigateurs, il sera considéré comme autant d'utilisateurs uniques. Les personnalisations de contenus adressées aux utilisateurs récurrents ignorent ainsi en réalité une masse considérable d'utilisateurs. Si un utilisateur se connecte à trois reprises à un même site avec trois appareils différents et finit par convertir, l'outil de tests A/B aura comptabilisé trois utilisateurs distincts pour une seule conversion.

Les outils de tests A/B collectent les informations côté client 33 et les envoient côté serveur sous la forme d'appels. Une requête HTTP est émise pour télécharger une image d'un pixel de haut sur un pixel de large. Les informations sont transmises sous la forme de paramètres puis traitées côté serveur. Il convient de distinguer trois déclencheurs à la transmission de ces informations:

- Le chargement d'une page.
- Le clic sur un élément prédéfini.
- La satisfaction d'une condition Javascript.

Si le premier et le troisième déclencheur ne posent pas de problèmes dans la plupart des cas, il en va tout autrement du second. Le déclenchement d'un événement au clic peut se faire au moment du clic proprement dit, c'est l'événement on Mouse Down, ou au moment où l'utilisateur relève son doigt. Dans ce dernier cas de figure il s'agit de l'événement onClick. La méthode employée varie selon les outils. Pourtant les écarts constatés peuvent être spectaculaires <sup>34</sup> et ceci en ignorant les possibles conflits entre une fonction préexistante et celle de l'outil de tests A/B. La conséquence ? Un risque de non transmission de l'information relative à l'action de l'utilisateur avec des gestions très diverses des conflits selon les navigateurs et leur version <sup>35</sup>. Sans oublier les informations non-transmises par la faute des outils d'aide au paramétrage proposés par les solutions elles même. Ainsi, il est possible de sélectionner n'importe quel élément présent sur la page dans l'éditeur WYSYWIG et de demander à suivre les clics l'ayant affecté. Pour se faire, l'outil repère l'emplacement de l'élément dans la page Web afin de définir un sélecteur. Or, les modifications appliquées dans le cadre d'un test sont de nature à

15

<sup>&</sup>lt;sup>32</sup> Fichier texte de taille réduite stocké dans le cache du navigateur.

<sup>&</sup>lt;sup>33</sup> Dans le cadre du modèle client-serveur, le terme client désigne les programmes s'exécutant sur la machine de l'utilisateur d'un service et le terme serveur, ceux hébergés et fonctionnant sur les serveurs de l'organisation proposant le service.

http://www.cardinalpath.com/experiment-onclick-vs-onmousedown-event-tracking-in-google-analytics/
 http://stackoverflow.com/questions/14048967/javascript-conflict-with-onclick-and-onmouseout

bouleverser en profondeur la structure du document HTML. Un sélecteur défini automatiquement est le plus souvent construit au moyen de règles trop précises pour demeurer efficace après la moindre modification. En cas de clic, aucun signal ne sera transmis sur les versions modifiées de la page, sans que cela s'avère détectable avant plusieurs jours.

Par ailleurs, la croyance dans le caractère irréfutable des chiffres occulte le rôle central joué par les algorithmes <sup>36</sup>. Ces derniers se voient confier la lourde tâche de déterminer si les résultats collectés peuvent être jugés viables d'un point de vue statistique. Or, il n'existe pas un algorithme unique utilisé par l'ensemble des solutions de tests A/B. Bien au contraire, les solutions mettent régulièrement à jour leur algorithme en insistant sur les progrès accomplis dans la réduction des erreurs et de la détection de faux positifs<sup>37</sup>. Ce faisant, elles passent sous silence les limites et erreurs attribuables à la précédente génération de leurs algorithmes. Comment expliquer autrement que par une défaillance des algorithmes la déclaration d'une variation gagnante en cas de *test A/A* <sup>38</sup>.

Les conséquences de la plupart de ces problèmes, à l'exception de celui affectant les algorithmes, pourraient être minorées si l'unité de mesure des solutions de tests A/B était non pas l'utilisateur, mais la session. Grâce à ce changement d'unité de mesure, les problèmes liés à la perte de l'identifiant utilisateur pourraient être minorés. Cependant, pour les sociétés commercialisant des solutions de tests A/B, ce serait reconnaître les limites de leur solution et nuire par la même à la croissance du marché des tests A/B. Une prise de risque inenvisageable dans un secteur aussi concurrentiel.

## Des palliatifs nombreux mais insatisfaisants

Les utilisateurs d'applications Web collectant des données côté client ont pendant des années cherché à trouver des palliatifs à la perte du cookie indispensable à l'identification des visiteurs. Le recours à la technologie flash de la société Adobe permit assez tôt d'offrir une solution au problème. En effet, le stockage et la sauvegarde des cookies flash fonctionnent différemment des cookies échangeables au moyen du protocole http. Afin de les supprimer, il s'avère nécessaire de les retirer de l'arborescence système en les plaçant dans la corbeille. La destruction nécessite donc de quitter le navigateur. La valeur des cookies flash réside dans leur résistance. Il s'agit également de leur principal point faible. Critiqués par les défenseurs de la vie privée sur le Web pour leur utilisation potentielle à des fins de pistage et de surveillance, les cookies Flash n'ont jamais été utilisés par des éditeurs de solution Web analytique. En revanche, des Web analystes ont pu recourir à cette technologie de façon ponctuelle pour stocker de façon pérenne des informations relatives à un utilisateur. Ils se servaient des cookies comme d'une base de données locale et injectaient leur contenu dans des cookies traditionnels. Dès lors, ces informations redevenaient exploitables par les outils de Web analytique. Outre le manque de contrôleur utilisateur sur la durée de vie des Flash Cookie, le caractère propriétaire de cette technologie a également fortement joué dans son rejet. Les luttes entre les géants américains du Web en vue d'imposer leurs propres standards technologiques au détriment de leurs concurrents ont également eu un rôle déterminant.

<sup>&</sup>lt;sup>36</sup> WANG TRICIA, *The Conceit of Oracles: How we ended of up in a world where quantitative data is more valued than qualitative data*, <a href="http://triciawang.com/updates/2014/2/5/transcript-of-my-talk-the-conceit-of-oracles-how-we-ended-of.html">http://triciawang.com/updates/2014/2/5/transcript-of-my-talk-the-conceit-of-oracles-how-we-ended-of.html</a>, février 2014.

<sup>&</sup>lt;sup>37</sup>http://blog.optimizely.com/2015/01/20/statistics-for-the-internet-age-the-story-behind-optimizelys-new-stats-engine/.

<sup>38</sup> http://blog.sumall.com/journal/optimizely-got-me-fired.html.

Apple fut le premier à lancer l'offensive contre Adobe en annonçant l'absence de support de la technologie Flash sur son système d'exploitation pour mobiles et tablettes <sup>39</sup>. Google, en position dominante sur le marché des Web analytique avec sa solution Google Universal Analytics, s'est joint plus tardivement à la lutte. Le dernier acte du bras de fer opposant les deux géants a été l'annonce de l'abandon progressif de Flash pour la diffusion des contenus vidéos sur *Youtube* <sup>40</sup> au profit d'un lecteur *HTML5*. Reste que ce sont principalement les possibilités offertes par la technologie HTML5 et son support par un nombre croissant de navigateur qui ont conduit à un recul de Flash. Privée de sa principale valeur ajoutée, la garantie de délivrer des fonctionnalités avancées quel que soit la version d'un navigateur, flash est devenu indésirable aux yeux des principaux acteurs du Web. Dernièrement, la fondation Mozilla a rendu publique une fonctionnalité expérimentale baptisée Shumway permettant de lire la technologie Flash sans avoir à utiliser un plugin <sup>41</sup>. Considérant ces tendances, les cookies Flash ne peuvent plus être considérés comme une voie de recours universelle au problème de l'absence de pérennité des cookies.

Tout comme les lecteurs HTML5 remplacent progressivement les lecteurs Flash sur les plateformes de diffusion de contenus audio et vidéo, le Local Storage est une technologie HTML5 ayant le potentiel de remplacer les cookies. Les cookies ne peuvent stocker que des quantités d'informations extrêmement limitées et sont transférées au serveur à chaque requête HTTP. A l'opposé, le local Storage permet de sauvegarder jusqu'à 5mb de données, sans nécessité de transmettre systématiquement ces informations aux serveurs. La valeur du local Storage réside dans la permanence des données déposées qui demeurent accessibles y compris après fermeture du navigateur <sup>42</sup>. Contrairement aux données flash, le local Storage peut être vidé comme les cookies à la demande de l'utilisateur. Il faut noter que le paramétrage par défaut des navigateurs exclue ce type de nettoyage des données. Plus durable que les cookies, le local storage n'en reste pas moins exposé à des risques d'effacement. En outre, il n'est pas supporté par les versions les plus anciennes d'Internet Explorer <sup>43</sup>. La valeur ajoutée de cette méthode sur les cookies ne s'avère pas suffisamment importante, notamment pour le stockage de faibles quantités d'informations.

En dépit de leurs défauts, les cookies restent plébiscités pour leur compatibilité avec l'intégralité du parc de navigateurs Web. S'ils n'ont pas été remplacés, c'est qu'aucune technologie alternative n'existe qui ne présente pas les mêmes faiblesses sans pour autant poser problème du point de vue du respect de la vie privée. Une méthode pour générer des cookies persistants baptisés « evercookie » a bien été mise au point. Néanmoins, elle s'avère tellement problématique d'un point de vue éthique qu'elle n'est pas utilisée par des solutions commerciales <sup>44</sup>. Ceci explique que les efforts d'une partie des éditeurs de solutions de Web analytique et de tests A/B se soient portés vers des moyens de régénérer à l'identique un cookie supprimé. Pour se faire, plusieurs solutions dont la suite de tests A/B et de Web Analytique Adobe Marketing Cloud ont recours à l'analyse de l'adresse IP <sup>45</sup>. Lorsque le code de la solution s'exécute, il vérifie en premier lieu la présence d'un cookie. En cas d'absence, il génère aléatoirement un identifiant unique et le stocke dans

<sup>&</sup>lt;sup>39</sup> https://wikileaks.org/wiki/Why\_Apple\_Won%27t\_Allow\_Adobe\_Flash\_on\_iPhone.

<sup>40</sup> http://www.theverge.com/2015/1/27/7926001/youtube-drops-flash-for-html5-video-default.

<sup>&</sup>lt;sup>41</sup> http://www.macg.co/logiciels/2015/02/firefox-commence-lire-le-contenu-flash-sans-le-plug-dadobe-87306.

<sup>42</sup> http://www.w3schools.com/html/html5\_webstorage.asp.

<sup>43</sup> http://caniuse.com/#search=storage.

<sup>44</sup> https://github.com/samyk/evercookie.

<sup>45</sup> http://blogs.adobe.com/digitalmarketing/analytics/want-a-cookie/.

un cookie. Lors du premier envoi d'informations aux serveurs de la solution, l'adresse IP du visiteur est transmise avec l'identifiant nouvellement créé. Une vérification de l'adresse IP est effectuée pour déterminer si un identifiant antérieur y était associé. Dans les cas où cette hypothèse se vérifie une correspondance est mise en place au moyen d'un système d'aliasing. Le dernier identifiant créé, bien que différent, permettra d'associer les données à un seul et unique utilisateur. Ce mode opératoire n'est cependant pas dénué de défauts :

- Il nécessite des investissements conséquents de la part de l'éditeur de la solution y ayant recours. En effet, la mise en place de cette fonctionnalité implique l'ajout d'une étape préliminaire de traitement des données avant leur stockage définitif dans des bases de données. Le stockage des adresses IP induit également des surcoûts.
- En Europe, la gestion, la conservation et l'exploitation des adresses IP est un sujet sensible. En France, la CNIL a restreint le droit de stockage aux seuls prestataires techniques 46. Les éditeurs de solutions se défendent de toute infraction à la loi dans la mesure où ils ne donnent pas l'accès à ces données à leurs clients.
- Enfin, une adresse IP ne peut être considérée comme un indicateur fiable :
  - o Du fait de son caractère changeant dans le cas d'une adresse IP dynamique.
  - O Du fait de son attribution potentielle à de nombreuses machines dans le cas d'une adresse IP fixe, notamment pour les machines hébergées au sein de structures collectives.
  - o Du fait de la possibilité d'utiliser un VPN pour masquer et modifier son adresse IP.

Du fait de ces nombreuses limites, cette méthode n'est exploitée qu'en dernier recourt par les solutions de tests A/B destinées exclusivement aux grands comptes tel qu'Adobe Marketing Cloud et Content Square.

S'il s'avère impossible d'identifier à coup sûr une machine au moyen de son adresse IP pour reconstituer un cookie, pourquoi ne pas essayer de faire de même mais avec le navigateur stockant le cookie ? Cette approche porte un nom : le browser fingerprinting <sup>47</sup>. Il repose sur l'analyse côté client de l'ensemble des éléments à même de distinguer un navigateur d'un autre parmi lesquels notamment :

- Le user-agent.
- La résolution de l'écran.
- Les polices installées.
- Les plugins installés.
- Les performances du navigateur pour le rendu d'éléments graphiques.

Si l'approche peut sembler séduisante, elle résiste mal à l'épreuve des faits. En effet, les éléments distinctifs ne sont pas suffisamment nombreux et divers pour éliminer le risque que deux navigateurs aient une empreinte similaire. Surtout, elle peine à convaincre du fait du caractère extrêmement fluctuant de certains éléments tels que la résolution de l'écran, les plugins installés et surtout la version du navigateur.

Le recours aux informations contenues dans un CRM apparaît comme une solution possible à cette impasse. En injectant dans le code source des pages Web des données

<sup>&</sup>lt;sup>46</sup> http://www.nextinpact.com/news/87162-au-conseil-detat-et-a-cnil-coup-frein-sur-collecte-adresses-ip.htm.
<sup>47</sup> https://w3c.github.io/fingerprinting-guidance/.

issues d'un CRM, il devient possible de recréer et d'alimenter convenablement un cookie, y compris en cas de suppression. Un inconvénient demeure : pour pouvoir délivrer les informations correspondant au visiteur, il est nécessaire qu'il soit connecté. Ce système apparaît approprié uniquement pour des sites Web auxquels l'accès est restreint par une demande de connexion <sup>48</sup>. Dans les faits, la connexion au CRM demeure inexploitable sur la majorité des sites Web.

Il n'existe pas à l'heure actuelle de solution efficace à même de concilier trois exigences contradictoires :

- Droit de regard de l'internaute sur les données stockées.
- Un stockage pérenne de l'information.
- Des technologies de stockage de l'information ou d'identification disponibles sur navigateur volatiles, limitées ou peu fiables.

## Contrôler les populations pour mieux tester ?

Quand bien même ces difficultés auraient été surmontées, un problème majeur demeurerait : la confusion entretenue par les éditeurs de solutions de tests A/B entre les termes navigateur et utilisateur. Les cookies sont stockés au sein d'un navigateur qui s'exécute sur une machine donnée. Un même utilisateur d'un service, entendre par là un être humain, peut être amené à utiliser divers navigateur. Par conséquent, il sera considéré comme autant d'utilisateurs distincts par les logiciels de tests A/B. L'ampleur du phénomène est encore aggravée par la croissance du nombre d'appareils disponibles au sein des foyers. Désormais, une part croissante d'utilisateurs d'un service Web ou d'un site Web utilisent plusieurs appareils ordinateurs, tablettes, smartphones pour y accéder avec pour conséquence la création à minima d'un utilisateur supplémentaire par appareil. A l'inverse, un même navigateur peut servir à plusieurs utilisateurs distincts sans aboutir à ce qu'un nouvel identifiant soit généré. La situation peut être aggravée par l'exploitation des données issues du CRM. En effet, l'utilisation d'un site ou d'un service Web par un tiers utilisant le navigateur d'un utilisateur demeuré identifié conduira à attribuer un comportement au mauvais individu.

Puisque l'unité de mesure des logiciels de tests A/B n'est pas l'individu, mais la machine, il ne peut y avoir qu'une piste d'améliorations envisageable afin de fiabiliser les données collectées : exercer un contrôle étroit sur les populations testées. L'ensemble des solutions disponibles sur le marché propose des critères de ciblage plus ou moins avancés selon le plan tarifaire choisi. Les critères disponibles non affectées par les biais induits par les cookies sont :

- la provenance géographique de l'internaute
- le canal d'arrivée sur le site
- la langue du navigateur
- le type d'appareil
- la résolution d'écran

Des critères extrêmement génériques qui ne donnent aucune indication sur des êtres humains mais uniquement des détails techniques sur les machines qu'ils utilisent. L'utilisation d'un CRM fournit davantage d'informations, mais sans qu'il soit possible de vérifier l'identité réelle de l'utilisateur d'un appareil sur lequel un visiteur s'est connecté. Le recrutement préalable à un test de personnes volontaires permettrait de

<sup>&</sup>lt;sup>48</sup> Le site vente-privee.com en est un bon exemple.

résoudre ce problème. En outre, elle aurait l'avantage de permettre d'obtenir davantage d'informations sur les utilisateurs testés que ne le permet une collecte automatisée de données.

Mais à quel prix ? La sélection des utilisateurs préalablement à l'exécution d'un test poserait en réalité davantage de problèmes qu'elle n'en résoudrait :

- Le comportement des utilisateurs serait altéré par leur connaissance du déroulement du test. Plus leur degré d'informations serait important, plus le test en serait biaisé.
- Le coût du recrutement d'une base d'utilisateurs prêts à subir un test pourrait à lui seul remettre en question la rentabilité d'une démarche de tests A/B.
- Au-delà des problématiques de coûts, les volumétries de visiteurs nécessaires à la réalisation de certains tests rendraient impossible la constitution d'un échantillon suffisant pour garantir la fiabilité statistique des résultats.
- Il s'avérerait impossible de quantifier le degré de représentativité du panel d'utilisateur testé.
- Quand bien même les visiteurs en viendraient à correspondre du point de vue de leur répartition socio-démographique aux visiteurs habituels, il serait impossible de s'assurer que leurs motivations s'avèrent similaires.

Un renforcement du contrôle sur l'identité des participants au test apparaît indispensable en vue de fiabiliser les données. Néanmoins cette approche s'avère impossible car elle remet en question les principes même des tests A/B, notamment l'exécution de tests en dehors de conditions réelles d'utilisation.

La fiabilité des données collectées lors de l'exécution d'un test A/B apparaît fortement diminuée par l'incapacité à comptabiliser l'identité et le nombre réel des utilisateurs. A l'heure actuelle, il n'existe pas de solutions techniques ou méthodologiques satisfaisantes pour résoudre efficacement ce problème, tout au plus des palliatifs <sup>49</sup>. Les problèmes identifiés au niveau des procédés de collecte des données et d'évaluation des performances conduisent à émettre de sérieuses réserves sur l'unité de mesure de référence des outils de tests A/B.

<sup>&</sup>lt;sup>49</sup> KAISER FUNG, *The reality is most A/B tests fail, and Facebook is here to help*, http://junkcharts.typepad.com/numbersruleyourworld/2014/04/the-reality-is-most-ab-tests-fail-and-facebook-is-here-to-help.html, avril 2014.

## Des données quantitatives pour principale source d'informations

Les difficultés d'identification des visiteurs inclus dans un test mettent en évidence les limites des processus de collecte de données passifs et automatiques. Pourtant, la confiance dans la valeur des tests A/B repose principalement sur la nature quantitative des données collectées lors d'un test. En ce domaine, les tests A/B ne constituent pas un phénomène isolé. En effet, la majorité des spécialistes des Big Data n'émettent pas de réserve quant à la fiabilité des informations les constituant <sup>50</sup>. Certains vont plus loin en prétendant que l'étude de ces ensembles de données permettrait de réduire les risques d'erreurs, de maximiser les retours sur investissement, voire de prédire l'avenir. Les thèses des défenseurs de l'exploitation des Big Data à des fins prédictives et décisionnelles postulent une valeur intrinsèque des données quantitatives <sup>51</sup>. La délégation du traitement des données et de leurs analyses à des algorithmes plutôt qu'à des êtres humains est justifiée avant tout par la quantité d'informations à traiter. Elle s'explique également par la place qu'occupent les statistiques dans les sociétés occidentales contemporaines. Dans un contexte de laïcisation et d'engouement sans précédent pour les possibilités offertes par l'électronique, le rapport aux données quantitatives revêt un caractère quasi mystique 52. Pourtant, l'étude des données quantitatives collectées lors de la réalisation d'un test A/B et de leur usage potentiel conduit à en minorer la valeur réelle.

## De nombreuses lacunes à surmonter

Pour s'en convaincre, il suffit de s'intéresser aux questions auxquelles se trouvent confrontées les équipes chargées du paramétrage de tests A/B et de l'analyse de leurs résultats. Le premier des problèmes qui se pose est le nombre d'indicateurs limités à disposition de l'analyste. Pour évaluer la performance d'une variation par rapport à la version d'origine, trois types d'objectifs sont définissables :

- La visualisation d'une page.
- Le clic sur un élément HTML.
- Le fait de remplir une condition évaluée par du Javascript. Les cas d'usage les plus courants sont :
  - o L'utilisation dans un environnement en AJAX 53.
  - o La comptabilisation du nombre total de pages consultées.
  - o L'interaction avec un élément sur la page.

La possibilité d'avoir recours au Javascript rend en théorie possible le suivi de tout type de comportement. Dans les faits, l'impossibilité d'appliquer des traitements rétroactifs aux données, hormis pour les pages vues, pose de graves problèmes en cas d'échec. En effet, le modèle de collecte et de traitement des données des solutions de tests A/B repose

<sup>&</sup>lt;sup>50</sup> Crawford Kate, *The Hidden Biases in Big Data*, <a href="https://hbr.org/2013/04/the-hidden-biases-in-big-data">https://hbr.org/2013/04/the-hidden-biases-in-big-data</a>, avril 2013.

<sup>&</sup>lt;sup>51</sup> Brooks Data, « What You'll Do Next », in : *The New York Times*, http://www.nytimes.com/2013/04/16/opinion/brooks-what-youll-do-next.html, avril 2013.

<sup>&</sup>lt;sup>52</sup> WANG TRICIA, *ibidem*.

<sup>&</sup>lt;sup>53</sup> Acronyme de Asynchronous JavaScript and XML. Il s'agit d'une architecture informatique permettant de construire des pages Web dynamiques sur le poste d'un client, sans jamais nécessiter d'opérer le moindre rechargement de la page Web.

sur l'externalisation des calculs côté client. Ce faisant, les sociétés éditrices de solutions de tests A/B réalisent d'importantes économies. A cet égard, la solution Content Square se distingue nettement de ses concurrents, dans la mesure où elle collecte systématiquement l'ensemble des clics et déplacements. La mise à disposition de cette fonctionnalité s'explique par le positionnement tarifaire élitiste de la solution qui permet de la financer. Ces limitations affectant les objectifs passent inaperçues tant que les indicateurs de performances d'un test évoluent positivement dans au moins l'une des variations. En revanche, en cas d'échec, les limites de cette approche apparaissent clairement. Spontanément, l'analyste va chercher des éléments explicatifs au maintien ou à la détérioration du niveau de performance des indicateurs :

- Les mauvaises performances affectent-elles l'ensemble des visiteurs testés ou uniquement une sous-population ?
- D'autres éléments et fonctionnalités ont-ils été affectés de façon similaire par le déploiement des modifications ?

Il pourrait répondre à la première question en utilisant les possibilités de segmentation de l'audience testée natives aux outils de tests A/B. Seulement les filtres disponibles sont limités en nombre. Pire encore, ils ne sont pas cumulatifs dans la plupart des plans tarifaires inférieurs des solutions disponibles sur le marché. Surtout, les combinaisons possibles s'avèrent tellement nombreuses que l'identification des segments d'audience les plus remarquables requiert l'aide d'un programme informatique. Quant à la seconde question, il s'avèrera impossible d'y répondre puisque les éléments et fonctionnalités n'étant pas directement concernés par le test n'auront pas fait l'objet d'un suivi. Les méthodes de définitions des indicateurs de performances des tests A/B ne donnent qu'une vision tronquée de la situation d'ensemble. Cette vue limitée peut potentiellement s'avérer trompeuse puisque les données quantitatives ne rendent compte que de l'évolution d'une série d'indicateurs limités. Ainsi, il se peut que le succès apparent d'un test ne soit en réalité qu'un échec 54. Sur un site E-commerce, la modification de l'emplacement et de la taille d'un bouton d'appel à l'action pourront permettre d'augmenter le pourcentage de visiteurs validant une commande. Néanmoins, l'augmentation du pourcentage d'acheteurs peut s'accompagner d'une diminution du pourcentage d'acheteurs satisfaits. De fait, la progression d'un indicateur peut s'accompagner de la régression d'un autre, sans qu'il soit possible de s'en rendre compte en utilisant uniquement un logiciel de tests A/B. Quand bien même la satisfaction des acheteurs aurait été identifiée comme l'un des objectifs du test, la collecte de données quantitatives n'aurait pas permis d'en suivre l'évolution. Non seulement la quantité de données quantitatives disponibles est considérablement limitée par le paramétrage effectué par l'utilisateur, mais en plus ces données ne permettent pas de suivre l'évolution de tous les types d'indicateurs.

A l'instar des données quantitatives qu'ils collectent, les tests A/B ne peuvent se suffire à eux même. Conscients que cette lacune est l'un des principaux points faibles des tests A/B, les éditeurs des solutions ont opté pour trois approches complémentaires :

 L'augmentation des possibilités de segmentation. Les filtres avancés demeurent cependant réservés aux souscripteurs des plans tarifaires les plus onéreux. Certaines sociétés, notamment Visual Website Optimizer, A/B Tasty et

<sup>&</sup>lt;sup>54</sup> NIELSEN JAKOB, *Ibidem*.

Optimizely, déclarent travailler sur des algorithmes permettant d'identifier automatiquement les typologies de visiteurs réagissant le mieux à une variation. Content Square propose d'ores et déjà un tel mécanisme. Cependant, la valeur des recommandations proposées est proportionnelle à la quantité d'indicateurs paramétrés.

- Sachant cela, l'ensemble des éditeurs de solutions offrent la possibilité de remonter des informations supplémentaires au sein de la plateforme de tests A/B grâce à une API. Il devient ainsi possible d'augmenter le nombre d'indicateurs disponibles et ce faisant d'élargir les possibilités de segmentations.
- Conscient de ces limites, du fait que l'étude du comportement utilisateur n'est pas l'apanage des seules solutions de tests A/B, que les utilisateurs des tests A/B utilisent simultanément ces outils complémentaires, qu'il n'existe pas de source de vérité unique sur le Web 55, les éditeurs insistent avant tout sur les possibilités de connexion de leurs outils avec des solutions tierces.

Ainsi, l'ensemble des solutions de tests A/B sur le marché proposent d'exploiter les APIs d'autres solutions pour augmenter la quantité d'informations disponibles. Le connecteur systématiquement proposé et le plus utilisé est sans conteste celui permettant d'interconnecter le logiciel de tests A/B à Google Analytics. Il devient ainsi possible d'exploiter l'ensemble des informations disponibles dans Google Analytics, dont l'exploitation et le paramétrage précèdent le plus souvent le recours à la solution de tests A/B, sans avoir à se soucier de les injecter dans l'outil de tests A/B. La connexion à des outils tiers permet en apparence d'avoir de la visibilité sur davantage d'indicateurs. En réalité, ce n'est qu'une illusion. En effet, le partage d'informations ne fonctionne que dans un sens : de l'outil de tests A/B vers l'outil tiers. Cependant, le périmètre d'intervention de l'algorithme implanté dans l'outil de tests A/B est restreint aux données stockées sur les serveurs de son éditeur. Il s'avère donc impossible de juger de la fiabilité statistique des tendances observées sur des données externes à un outil de tests A/B.

Les données quantitatives collectées par les solutions de tests A/B ou des outils tiers ne permettent pas de suivre de façon exhaustive et satisfaisante la performance d'une variation par rapport à la version d'origine. Le fonctionnement des outils est autant responsable de cet état de fait que les biais inhérents aux données quantitatives. Dans ce contexte, qualifier les tests A/B d'objectifs et d'irréfutables semble inapproprié <sup>56</sup>. Pourtant, il s'agit de leurs attributs les plus couramment invoqués pour justifier leur emploi. Pour contourner ce manque de valeur intrinsèque des données quantitatives, les spécialistes ont recours aux données qualitatives.

<sup>&</sup>lt;sup>55</sup> KAUSHIK AVINASH, *Best Web Analytics Tools: Quantitative, Qualitative, Life Saving!*, http://www.kaushik.net/avinash/best-web-analytics-tools-quantitative-qualitative/, 19 octobre 2010.

<sup>&</sup>lt;sup>56</sup> KOWITZ BRADEN, Should designers trust their instincts — or the data?, <a href="http://www.gv.com/lib/design-instinct-vs-data">http://www.gv.com/lib/design-instinct-vs-data</a>, 15 janvier 2014.

En effet, si les tests A/B permettent de savoir imparfaitement comment se comportent les visiteurs d'un site Web, ils s'avèrent totalement inadaptés à qui voudrait savoir pourquoi <sup>57</sup>. Afin de pouvoir comprendre la cause d'un échec et ainsi éviter de le reproduire, ou pour identifier les raisons d'un succès pour en tirer le meilleur parti, les spécialistes prônent le recours aux données qualitatives <sup>58</sup>. En collectant des données qualitatives avant, pendant et après la réalisation d'un test A/B, il serait ainsi possible de limiter les risques d'échecs et de maximiser le retour sur investissement <sup>59</sup>. Pourtant, les éditeurs de tests A/B mettent en avant les défauts des données qualitatives comme autant de raisons de les rejeter en faveur des tests A/B <sup>60</sup>. Les lacunes qui leurs sont le plus couramment reprochées sont notamment :

- 1. L'absence de visibilité sur la représentativité de l'échantillon interrogé et des réponses collectées.
- 2. Le coût prohibitif de la collecte et de l'analyse des données qualitatives.
- 3. Une analyse des données qui reste impossible pour le plus grand nombre.
- 4. La taille limitée des populations de répondants.
- 5. La subjectivité des réponses des répondants qui sont souvent en contradiction avec leurs actes.
- 6. La non prise en compte du point de vue des personnes ayant choisi de ne pas participer. <sup>61</sup>

Ces remarques ne résistent pas à une analyse sérieuse et apparaissent davantage comme des jugements subjectifs, qui sont la conséquence de préjugés négatifs envers les données qualitatives <sup>62</sup>. Les a priori concernant les données qualitatives s'expliquent pour l'essentiel par leur association avec les méthodes de collecte qui leur sont traditionnellement associées : l'entretien utilisateur et le sondage direct. En élargissant le champ des processus de collecte de l'information, ces critiques perdent de leur pertinence :

 L'absence de représentativité de l'échantillon des répondants est discutable dans la mesure où il s'avère possible de le segmenter grâce à une collecte passive d'informations et/ou de façon plus explicite grâce à des questions. Dès lors il devient possible d'effectuer des comparaisons avec la population de visiteurs habituels.

<sup>&</sup>lt;sup>57</sup> BERNHEIM BRUSH A.J., *Ibidem*.

<sup>&</sup>lt;sup>58</sup> PATEL NEIL, What Spending \$252,000 On Conversion Rate Optimization Taught Me, http://www.quicksprout.com/2012/10/08/what-spending-252000-on-conversion-rate-optimization-taught-me/, 8 octobre 2012.

CRAWFORD KATE, Ibidem.

<sup>&</sup>lt;sup>59</sup> KOWITZ BRADEN, *Ibidem*.

<sup>&</sup>lt;sup>60</sup> L'usage combiné des données qualitatives et quantitatives est pourtant à la base du Web analytics 2.0. En voici une définition : "the analysis of qualitative and quantitative data from your website and the competition, to drive a continual improvement of the online experience of your customers and prospects, which translates into your desired outcomes (online and offline)." Source : KAUSHIK AVINASH, *Ibidem*.

<sup>&</sup>lt;sup>61</sup> Pour une analyse plus poussée des différentes lacunes des données qualitatives et des méthodologies de collecte leur étant associées, voir : GOWARD CHRIS, *You should test that !*, John Wiley & Sons, Indianapolis, 2013, pp. 33-38. <sup>62</sup> Pour une approche plus subtile visant à dépasser la vision de données qualitatives et quantitatives intrinsèquement opposées, voir : LAI MICHAEL, *When Does Quantity Become Quality? How to navigate big data*, <a href="http://uxmag.com/articles/when-does-quantity-become-quality-how-to-navigate-big-data">http://uxmag.com/articles/when-does-quantity-become-quality-how-to-navigate-big-data</a>, 23 septembre 2013.

- 2. Les coûts prohibitifs induits par la collecte de données peuvent être considérablement réduits par une diffusion des sondages en ligne <sup>63</sup>. Les réponses subissent automatiquement un traitement préparatoire et sont disponibles directement sous forme numérique. L'analyse des données proprement dite constitue la seconde cause d'augmentation des coûts car elle s'avère extrêmement chronophage. Le laps de temps qui leur est consacré peut être réduit grâce à des traitements informatiques.
- 3. En ce qui concerne la plus grande complexité qu'il y aurait à baser une analyse sur des données qualitatives plutôt que sur des données quantitatives : du fait des biais multiples qui affectent les données quantitatives, dont les plus importants ont été précédemment évoqués, leur analyse nécessite non seulement de bonnes connaissances mais aussi une solide expérience. La capacité à porter un regard critique sur les données quantitatives ne va pas de soi. Il est vrai que les experts obtiennent de meilleurs résultats que des débutants lorsqu'ils mènent une étude qualitative. Cependant, dans le cas d'une étude quantitative, seuls les meilleurs d'entre eux sont à même d'arriver à des résultats valides 64. La supposée accessibilité supérieure des données quantitatives repose sur la capacité du plus grand nombre à lire les nombres qui font immédiatement sens. A l'inverse, les contenus textuels, qui constituent le matériau de base des données qualitatives, peut apparaître plus complexe à appréhender de prime abord <sup>65</sup>. Cependant, la simplicité des chiffres constitue également leur principal point faible. Leur capacité à synthétiser ou matérialiser des notions complexes les rend ambigus par nature. Dans le cadre d'un test A/B, le nombre de conversions associées à un objectif montre combien l'apparente accessibilité des données quantitatives est en réalité trompeuse. En effet, pour bien comprendre les données qui lui sont présentées, l'utilisateur doit savoir que les objectifs sont dédupliqués 66 sans quoi il commettra de lourdes erreurs d'analyse. 67
- 4. Le problème de la taille des populations de répondants est probablement le plus couramment évoqué. Il renvoie aux sessions d'entretiens utilisateurs contrôlées, en petit comité, excédant rarement plus d'une dizaine de participants. Ces limitations appartiennent au passé. Grâce aux questionnaires et micro-sondages en ligne, au chat et à la visio-conférence il est aujourd'hui possible de cibler l'ensemble des utilisateurs potentiels d'un site Web. Cependant, poser une question en ligne revient le plus souvent à ne pas obtenir de réponse. En effet, contrairement aux entretiens utilisateurs classiques qui s'effectuaient le plus

<sup>63</sup> James Kyle, *Are you ready for Web Analytics 2.0?*, <a href="http://doteduguru.com/id245-introducing-web-analytics-2-0.html">http://doteduguru.com/id245-introducing-web-analytics-2-0.html</a>, juin 2008.

<sup>&</sup>lt;sup>64</sup> NIELSEN JAKOB, *Ibidem*.

<sup>&</sup>lt;sup>65</sup> NIELSEN JAKOB, *Ibidem*.

<sup>66</sup> Lorsqu'un même utilisateur réalise à plusieurs reprises une action ayant été définie comme un objectif, par exemple le clic sur un bouton, seule la première des conversions est comptabilisée par l'outil. La notion de volume n'apparaît pas dans les résultats. La déduplication est utilisée par l'ensemble des outils de tests A/B du marché. Le recours à cette méthode de comptabilisation répond à un impératif de fiabilisation des données. En effet, dans la majorité des cas le volume d'individus ayant réalisé une action compte davantage que le volume d'actions réalisées. Cette approche a cependant ces limites dans des cas de figure où le volume d'individus compte autant que le volume d'actions, par exemple dans le cadre d'un test portant sur un moteur de recherche. En effet, connaître le nombre de recherches effectué par un internaute est tout aussi important que de connaître la part des visiteurs ayant utilisé le moteur de recherche. Puisqu'un outil de tests A/B ne permet pas d'effectuer un suivi de la fréquence de recherche, il s'avère nécessaire d'utiliser un outil tiers pour lequel l'analyse de la fiabilité statistique des résultats ne sera pas disponible.

<sup>&</sup>lt;sup>67</sup> Pour une approche critique des méthodes de collecte de données qualitatives en ligne, voir : GATTIKER E. URS, *Why A/B testing of web design fails*, <a href="http://commetrics.com/articles/supplement-with-qualitative-assessment/">http://commetrics.com/articles/supplement-with-qualitative-assessment/</a>, décembre 2009.

souvent contre rémunération, la participation en ligne est totalement désintéressée et repose sur le volontariat. Payer les répondants risquait certes de les influencer, mais permettait d'obtenir des taux de réponse de 100%. En ligne, la participation des utilisateurs varie fortement en fonction de la méthode de collecte, de la nature ouverte ou fermée des questions, de leur nombre et du contexte dans lequel les internautes sont sollicités. Avec des pourcentages de participation pouvant osciller de moins d'un pourcent à plusieurs dizaines, il est impossible d'avancer un taux de réponse moyen qui ait du sens. Reste qu'avec des volumétries de visiteurs mensuels atteignant à peine quelques milliers et un taux de réponse de 1%, il s'avère déjà possible de toucher un nombre considérable d'individus sur la durée.

- 5. Le décalage entre le ressenti d'un utilisateur et les données comportementales collectées passivement peut s'expliquer par l'effet déformant de la mémoire, notamment lorsque les retours utilisateurs sont collectés en dehors du contexte d'utilisation habituel du dispositif. Lorsque l'utilisateur est interrogé pendant ou immédiatement après sa session et que des décalages apparaissent entre ses déclarations et ses actes, c'est que le contexte dans lequel les données sont collectées, la façon dont sont formulées les questions et le niveau de stress ont pesé sur ses réponses. Sans nier l'existence de ces biais, il faut voir dans les écarts entre les retours utilisateurs et leurs actions davantage une opportunité qu'un problème.
- 6. Comment défendre cette approche alors que l'avis des personnes qui choisissent de ne pas répondre est ignoré puisque totalement absent des données analysées ? Au-delà des différents éléments qui conduisent à minorer les critiques adressées aux données qualitatives, ce sont les causes de leur emploi qui permettent de déterminer leur valeur pour une démarche donnée. Dans le cas d'une démarche de tests A/B, le rôle dévolu aux données qualitatives est triple :
  - Fournir à l'analyste des hypothèses et des éléments à même de nourrir sa réflexion dans le cadre de la conception d'un test A/B.
  - Guider l'analyste dans son analyse des données quantitatives recueillies au cours du test afin de limiter les risques de mauvaise interprétation des données.
  - Permettre d'expliquer la réussite ou l'échec d'un test.

Les critiques adressées aux données qualitatives n'ont pas lieu d'être dans le cadre de la réalisation des tests A/B, car elles jouent un rôle complémentaire à celui des données quantitatives qu'elles complètent. Lors de la réalisation de tests A/B, les données qualitatives ne constituent jamais un indicateur de performance à même de décider de la réussite ou de l'échec d'un test. Dès lors, qu'importe que la majorité des visiteurs n'aient pas répondu aux questions posées <sup>68</sup>, ou encore le décalage entre leurs déclarations et les données comportementales puisque les faiblesses des données qualitatives s'avèrent compensées par les données quantitatives et inversement. La comparaison d'ensemble de

Cette approche sous-tend que les données sélectionnées sont représentatives de celles exclues de l'analyse. Si ce présupposé était transposé aux données qualitatives il suffirait à écarter tout doute quant à leur fiabilité.

<sup>&</sup>lt;sup>68</sup> La pratique de l'échantillonnage sur les données quantitatives est pourtant une pratique courante qui ne suscite pas un tel rejet. Etant donné les volumes de données massifs collectés par les outils de Web Analytiques la plupart d'entre eux réservent les analyses sur des ensembles de données exhaustifs à la souscription d'offres tarifaires accessibles uniquement aux grands comptes. Concrètement, les éditeurs de ces solutions procèdent à un double échantillonnage :

<sup>•</sup> En amont, le volume de données transmissibles quotidiennement est soumis à un seuil au niveau du site Web tout entier et/ou au niveau du visiteur.

Lorsque la base de données de l'outil de Web analytiques est interrogée pour analyse, une fraction de l'ensemble des données est choisie aléatoirement. Ensuite, le logiciel extrapole sur la base de ces seules données des résultats applicables à l'ensemble du site.

données d'origine et de nature distinctes constitue la réponse des spécialistes des tests A/B aux lacunes des outils de test qu'ils utilisent. Les spécialistes proposent une méthodologie qui fait s'entremêler à plusieurs reprises les données qualitatives et quantitatives. En voici les principales étapes :

- 1. Mener un travail de réflexion sur les objectifs stratégiques de son site Web. Identifier les indicateurs opérationnels quantifiables permettant de rendre compte de la performance du dispositif vis-à-vis des objectifs stratégiques.
- 2. Identifier les pistes d'optimisation existantes les plus intéressantes afin de faire progresser les indicateurs opérationnels identifiés à l'étape 1. Les logiciels de Web Analytique permettent notamment de détecter les points hémorragiques dans la navigation tels que les pages présentant un *taux de rebond* <sup>69</sup> conséquent ou une *valeur de page* <sup>70</sup> peu élevée ou encore un *taux de sortie* <sup>71</sup> important. Le paramétrage de *cartes de chaleur* <sup>72</sup>, le visionnage d'*enregistrements de sessions utilisateurs* <sup>73</sup> conjugués à la diffusion de sondages et micro-sondages en ligne permettent d'obtenir des données complémentaires à celles des outils de Web analytique et d'aider l'analyste à les interpréter.
- 3. Paramétrer un test A/B ou multivarié en tenant compte des données collectées à l'étape 2 dans le but de faire progresser les indicateurs définis à l'étape 1.
- 4. Lancer le test, exécuter simultanément des relevés de cartes de chaleur ainsi que des micro-sondages en ligne afin de récolter un maximum d'informations utiles à l'issue du test lors de l'analyse des résultats.
- 5. Effectuer un bilan du test, les données collectées par les outils de Web analytique, de cartes de chaleur et de micro-sondages doivent permettre de ne pas commettre d'erreurs d'interprétation et d'expliquer la réussite ou l'échec d'un test.

Si cette méthodologie permet d'augmenter le taux de réussite d'un test et de maximiser les résultats obtenus par une variation comparée à l'original, ce qui explique que les professionnels déconseillent fortement d'effectuer des tests A/B sans l'appliquer, elle présente également plusieurs inconvénients majeurs :

• Son coût prohibitif tout d'abord, puisqu'elle nécessite de recourir au minimum à cinq type d'outils distincts : Web analytique, Sondages, Micro-sondages, Cartes de chaleur, tests A/B. Au-delà des questions d'acquisition des licences qui peuvent conduire à augmenter les coûts, le paramétrage des outils présuppose un travail conséquent en amont. Pour les outils de Web analytique, il s'agira de disposer d'un plan de marquage à jour qui puisse garantir la validité des données collectées. Pour les outils de carte de chaleur, de sondage et de micro-sondages, leur utilisation nécessite l'inclusion de leur code dans celui des pages. En outre, l'élaboration des questions, des règles de ciblage des pages et le paramétrage de

<sup>&</sup>lt;sup>69</sup> Le taux de rebond correspond au pourcentage de sessions ayant débuté sur une page ou un ensemble de pages au cours desquelles l'utilisateur a consulté moins de deux pages.

<sup>&</sup>lt;sup>70</sup> La définition d'objectifs, auxquels est associée une valeur, est une étape incontournable du paramétrage d'un outil de Web analytiques. Lorsqu'un visiteur atteint un objectif au cours d'une session, une partie de la valeur générée est automatiquement attribuée à l'ensemble des pages consultées au cours de cette même session. Si l'indicateur de valeur d'une page est faible, ceci signifie qu'elle ne constitue pas une étape incontournable du processus de conversion.

<sup>&</sup>lt;sup>71</sup> Part d'une page ou d'un ensemble de pages dans l'ensemble des pages ayant constitué la dernière page vue d'un internaute au cours d'une session.

<sup>&</sup>lt;sup>72</sup> Outil permettant de connaître les zones les plus cliquées, la profondeur de défilement et les déplacements de souris sur une ou plusieurs pages Web. Selon les plans tarifaires et les solutions, il est possible de segmenter les données obtenues par langue, provenance géographique, type d'appareil, etc.

<sup>&</sup>lt;sup>73</sup> Certains outils de Web Analytique et la plupart des outils de cartes de chaleur proposent d'enregistrer l'ensemble d'une session utilisateur, ce qui inclue les déplacements de la souris les clics et même les textes saisis, dans le but d'identifier après les avoir consultées des opportunités d'optimisation.

l'outil prennent également du temps. L'interfaçage des différents outils, afin de permettre une lecture des résultats par variation n'est pas toujours supporté de façon native et peut nécessiter des développements. Enfin, après avoir mis en place la collecte, il faut encore consacrer du temps à l'analyse de trois à quatre source de données distinctes à chaque phase qu'il s'agit de recouper.

- Son caractère élitiste. La maîtrise de cinq solutions distinctes, tant dans leur paramétrage que dans l'analyse des informations collectées, suppose de faire appel à un profil expert ou à une équipe qui maitrise les différents outils. La promesse de tests A/B peu onéreux, accessibles à tous et ne nécessitant pas de compétence particulière <sup>74</sup> pour être lancés apparaît totalement irréaliste.
- Son manque de rigueur. Si la cause de la multiplication des sources de données est parfaitement justifiable, la comparaison d'ensembles de données aussi disparates apparaît douteuse. En effet, les logiciels de Web Analytique partagent la plupart des limitations techniques des solutions de tests A/B auxquelles vient s'ajouter la question de l'échantillonnage des informations collectées. Les cartes de chaleur sont difficilement exploitables sur des ensembles de pages, y compris ceux partageant des gabarits similaires du fait de différences de profondeur et d'emplacements de certains contenus d'une page à une autre. En outre, les solutions disponibles sont incapables de représenter les éléments créés de façon dynamique après le chargement de la page. Les enregistrements de sessions utilisateurs pour être pleinement exploitables supposent l'enregistrement des caractères saisis. Or, ceci conduit à stocker chez un tiers des données personnelles sans accorder à leurs propriétaires un droit d'accès ou de rectification, ce qui est en contradiction avec les directives de la CNIL 75. Il en va de même avec les sondages et micro-sondages. Au-delà des défauts inhérents à chaque typologie d'outils, le véritable problème réside dans le recoupement entre ces ensembles de données.

## La comparaison:

- de données Web Analytique échantillonnées.
- de résultats de tests A/B à la fiabilité discutable.
- de quelques enregistrements de sessions utilisateurs sélectionnés empiriquement
- de cartes de chaleur imprécises.
- de réponses à des questionnaires et des micro-sondages apportées par une partie seulement des visiteurs.

fait-t-elle sens?

http://en.beampulse.com/ab-testing/.
 http://www.cnil.fr/index.php?id=20.

La comparaison d'ensemble de données divers ne suffit pas à répondre aux impératifs de rigueur et de fiabilité qui accompagnent la réalisation de tests A/B. Afin de circonscrire les défauts de chacun des outils et tirer le meilleur parti des tests A/B, la fusion des ensembles de données apparaît comme une solution possible. Elle nécessite

d'attribuer à chaque visiteur un identifiant unique qui sera associé aux données collectées par chacun des outils. S'il s'avère pour 1'heure impossible d'attribuer un identifiant pérenne, navigateurs et multi-appareils aux visiteurs non identifiés, le recours à un identifiant spécifique à un navigateur, généré aléatoirement, stocké dans un cookie et sujet à l'effacement constitue malgré tout une base de travail intéressante pour plusieurs raisons:

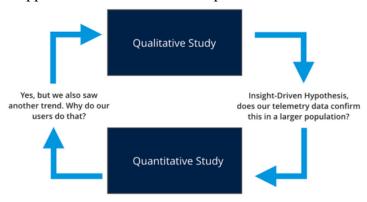


Fig. 2 : Modèle relationnel illustrant l'usage combiné des études qualitatives et quantitatives

Source: Selman Bill, Why Do We Conduct Qualitative User Research? https://blog.mozilla.org/ux/2014/10/why-do-we-conduct-qualitativeuser-research/

- Il s'agit du système d'authentification que partagent les outils de tests A/B, de Web Analytique, de sondages, de micro-sondages et de cartes de chaleur.
- Tous ces outils sont dotés d'APIs qui proposent soit la réécriture du cookie, soit la transmission d'informations additionnelles lors de l'envoi d'un paquet de données aux serveurs.

En associant aux données collectées par chaque solution l'identifiant d'un visiteur, il deviendrait possible de véritablement réconcilier les données. Une fois le problème posé par la représentativité intrinsèque de chacun des ensembles de données écarté, le véritable enjeu posé par la comparaison d'ensembles de données multiples et disparates apparaît : comment s'assurer que le recoupement de plusieurs sources de données s'accompagne d'un recoupement des utilisateurs ?

Dans le cadre de la préparation d'un test A/B, la consultation des données Web Analytique indique que la page d'accueil présente un taux de rebond anormalement élevé. Par ailleurs, un micro-sondage exécuté sur cette page, ainsi qu'une question d'un sondage ont permis de collecter des suggestions pour l'améliorer. Cependant, rien ne permet d'affirmer que les données qualitatives collectées l'ont été exclusivement auprès de visiteurs ayant rebondi. Quand bien même cela aurait été le cas, il se peut que les aspirations des visiteurs n'ayant pas rebondi s'opposent à celle des visiteurs ayant rebondi. Lancer le test sur la seule base de cette comparaison revient à faire un double pari :

- Parmi les répondants se trouvaient les visiteurs dont le comportement a été identifié dans l'outil de Web analytique.
- Les aspirations de cette frange de visiteurs ne sont pas opposées à celles des autres visiteurs.

Grâce à la fusion des sources de données, l'identification d'un segment de visiteurs dans l'outil de Web analytique permet d'exporter les identifiants qui leurs sont associées. Dès lors, il suffit de restreindre l'analyse aux réponses comportant ces identifiants pour limiter

les risques d'erreur d'interprétation <sup>76</sup>. L'identifiant généré par la solution de Web analytique doit servir de référence dans la mesure où l'utilisation de ce type de logiciel précède généralement celle des tests A/B. En outre, ils collectent des informations en continu et sur l'ensemble du site, tandis que les autres outils sont utilisés ponctuellement et sur un périmètre restreint. En outre, la fusion des ensembles de données rend théoriquement possible le rapprochement entre des données comportementales collectées plusieurs mois avant l'acquisition de données qualitatives.

La fusion des ensembles de données n'est pas une « solution miracle » puisqu'elle ne supprime pas le problème posé par l'identification des visiteurs, ni celui de l'inflation des coûts induits par la multiplication des outils et des analyses. Elle permet en revanche de limiter considérablement les erreurs d'analyse et les extrapolations abusives 77. Elle constitue un garde-fou indispensable à tout analyste qui souhaiterait recourir à une multiplication des types et sources de données pour fiabiliser les résultats obtenus grâce aux tests A/B. La fusion des sources de données permet également de répondre à la problématique de contrôle accru des visiteurs testés afin de contourner les lacunes techniques des outils de tests A/B. Plutôt que de limiter en amont du test les personnes incluses, la fusion des données permet de restreindre l'analyse aux seuls profils jugés fiables. Seuls les visiteurs pour lesquels plusieurs sources de données seraient disponibles se verraient inclus dans une analyse, avec pour conséquence un allongement de la durée des tests, pour atteindre des quotas de visiteurs « documentés » suffisamment importants. Une solution alternative à la fusion des ensembles de données, plus simple à mettre en œuvre et nécessitant des compétences et des investissements moindres, existe. Elle permettrait également de diminuer les risques d'erreurs, de maximiser les résultats et

d'expliquer les échecs et les réussites. Il s'agit des entretiens utilisateurs <sup>78</sup>. Evoquer la possibilité de réaliser des entretiens utilisateurs afin d'améliorer les résultats d'un test A/B, alors même que ces derniers se positionnent comme une alternative plus crédible à cette méthode traditionnelle d'étude l'expérience utilisateur sembler provocateur. Cependant, au demeurant, ces méthodes apparaissent davantage complémentaires <sup>79</sup>. Le manque qu'opposées crédibilité des entretiens utilisateurs aux yeux des personnes extérieures domaines l'expérience de

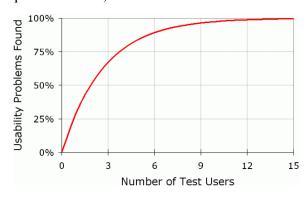


Fig. 3: Rapport entre le nombre d'utilisateurs testés et le nombre de problèmes ergonomiques détectés

Source: Why you only need to test with 5 users

<a href="http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/">http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/</a>

utilisateur et de l'ergonomie est difficile voire impossible à surmonter. Cependant, si les entretiens utilisateurs sont présentés comme une méthode parmi d'autres visant à collecter

<sup>77</sup> Alhadi Farid, *Combining Qualitative and Quantitative Data to Gain Useful Insights*, <a href="http://www.e-nor.com/blog/web-analytics/combining-qualitative-and-quantitative-data-to-gain-useful-insights">http://www.e-nor.com/blog/web-analytics/combining-qualitative-and-quantitative-data-to-gain-useful-insights</a>, 8 mars 2013.

<sup>78</sup> USERLYTICS, *Optimizing A/B User Testing*, <a href="http://www.userlytics.com/blog/optimizing-ab-testing">http://www.userlytics.com/blog/optimizing-ab-testing</a>, 30 septembre 2011.

30

<sup>&</sup>lt;sup>76</sup> DRISCOLL DAVID L.; APPIAH-YEBOAH AFUA; SALIB PHILIP; RUPERT DOUGLAS J., « Merging Qualitative and Quantitative Data », in: *Mixed Methods Research: How To and Why Not*, University of Georgia, vol. 18, 2007, <a href="http://digitalcommons.unl.edu/icwdmeea/18">http://digitalcommons.unl.edu/icwdmeea/18</a>.

 $<sup>^{79}</sup>$  Selman Bill, Why Do We Conduct Qualitative User Research?, <a href="https://blog.mozilla.org/ux/2014/10/why-do-we-conduct-qualitative-user-research">https://blog.mozilla.org/ux/2014/10/why-do-we-conduct-qualitative-user-research</a>/, 30 octobre 2014.

des pistes d'optimisation et à expliquer les résultats d'un test, leur usage ne posera pas davantage de problèmes que le recours aux sondages et micro-sondages <sup>80</sup>. Or, la complexité de mise en œuvre et le retour sur investissement potentiel des entretiens utilisateurs dépassent largement ceux des méthodes complémentaires aux tests A/B

#1: Find Your Best Opportunities

#2: Understand Visitor Needs

#3: Use Data & Insights to Inform Testing

#4: Where to Start

#5: Measuring Success

Fig. 4 : Un exemple de méthodologie pour conduire efficacement des test A/B

Source: How to Build a Strong A/B Testing
Plan That Gets Results
<a href="http://conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results">http://conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results</a>.

évoquées jusqu'à présent, à l'exception des outils de Web analytique. Grâce aux effets d'échelles, il devient possible de détecter plus de 80% des problèmes ergonomiques en ne sollicitant pas plus de cinq utilisateurs. Dès lors, les tests utilisateurs pourraient jouer un rôle déterminant dans un processus de tests A/B comme suit :

- 1. Mener un travail de réflexion sur les objectifs stratégiques de son site Web. Identifier les indicateurs opérationnels quantifiables permettant de rendre compte de la performance du dispositif vis-à-vis des objectifs stratégiques.
- 2. Identifier les pages, éléments de navigation ou fonctionnalités semblant offrir les meilleures opportunités d'amélioration grâce à l'outil de Web

analytique. Créer un *segment* <sup>81</sup> dans l'outil de Web Analytique de visiteurs dont le comportement indique l'existence d'un potentiel d'amélioration du site.

- 3. Récupérer les identifiants de ces visiteurs. Proposer aux visiteurs dont l'identifiant correspond à ceux exportés de participer à une session de test à distance. Mener les entretiens, collecter les données.
- 4. Paramétrer et lancer le test.
- 5. En cas d'échec ou de succès mener plusieurs entretiens, auprès de personnes ayant permis de tester l'une des variations, afin d'en déterminer les causes. <sup>82</sup>

Cette démarche requiert l'installation d'un seul outil supplémentaire sur le site : celui permettant de recruter des utilisateurs pour un test. Contrairement aux autres outils de collecte, les tests peuvent être menés rapidement et ne nécessitent aucun paramétrage. Ils offrent davantage de souplesse et d'agilité et s'avèrent moins chronophages et par conséquent moins onéreux.

<sup>81</sup> Un segment est une fonctionnalité permettant de limiter l'affichage des données sur la base de critères multiples et cumulatifs. Exemple : afficher uniquement les données des visiteurs mobiles dont le navigateur était paramétré en français et qui visitaient le site à 4 heures de l'après-midi.

Young Scott W. H., *Improving Library User Experience with A/B Testing: Principles and Process*, <a href="http://quod.lib.umich.edu/w/weave/12535642.0001.101?view=text;rgn=main">http://quod.lib.umich.edu/w/weave/12535642.0001.101?view=text;rgn=main</a>, Montana State University, Volume 1, Issue 1, 2014.

<sup>&</sup>lt;sup>80</sup> Ma Shanshan, *Pairing Up Usability Testing with A/B Testing*, http://www.uxmatters.com/mt/archives/2011/07/pairing-up-usability-testing-with-ab-testing.php, juillet 2011.

<sup>&</sup>lt;sup>82</sup> Cette méthodologie est proposée avant tout pour illustrer à quel stade le recours aux entretiens utilisateurs pourrait s'avérer le plus pertinent. Il n'existe pas un mais de multiples plans de test proposés par les professionnels afin de réaliser efficacement un test A/B. Nombreux sont ceux qui ne tiennent pas compte de la plupart des lacunes évoquées dans ce travail. La proposition de plan de test de Will Kurt se détache dans la mesure où elle est accompagnée d'une énonciation des risques d'erreurs courants dans le cadre de la réalisation de tests A/B. Pour en savoir plus, voir : Kurt Will, Most of Your A/B Test Results Are Illusory and That's Okay, <a href="https://blog.kissmetrics.com/your-ab-tests-are-illusory">https://blog.kissmetrics.com/your-ab-tests-are-illusory</a>, avril 2014.

La conduite d'entretiens utilisateurs, plus encore que la fusion de données de sources et de types divers, permet de pallier aux limites d'une approche purement comparative en ayant une connaissance plus poussée des caractéristiques des individus évalués <sup>83</sup>.

Les outils de tests A/B se contentent de collecter des données quantitatives. Ce parti pris, conjugué aux lacunes des méthodes de collecte d'informations des solutions présentes sur le marché, impose de leur adjoindre des sources de données extérieures. Pour autant, la multiplication des sources et types de données ne constitue pas à elle seule une réponse adaptée. Le croisement des ensembles de données apparaît indispensable en vue d'atteindre une plus-value qualitative. L'augmentation du nombre d'outils utilisés, en accroissant de façon exponentielle le temps à consacrer aux paramétrages puis à l'analyse de données incitent à limiter le recours à la fusion des ensembles de données. En revanche, l'utilisation d'un outil de Web analytique, conjointement à la réalisation d'entretiens utilisateurs à distance, semble à même d'offrir un excellent rapport coût/retour sur investissement.

L'évolution spectaculaire des logiciels de tests A/B a rendu leur usage accessible au plus grand nombre grâce à des interfaces soignées et des plans tarifaires n'excédant pas quelques centaines d'euros par mois. Néanmoins, cette accessibilité de façade cache des limitations techniques multiples qui rendent impossible une analyse basée sur les seules données des outils de tests A/B. Tout comme les données quantitatives, les tests A/B ne se suffisent pas à eux même. Leur usage suppose la maîtrise d'une méthodologie complexe et précise. Au-delà des problèmes soulevés par les limitations techniques des outils, qu'en est-il de la démarche proprement dite ?

<sup>&</sup>lt;sup>83</sup> Innes Jon, *Usability Testing Is Qualitative Only If You Can't Count*, <a href="http://www.uxmatters.com/mt/archives/2011/02/usability-testing-is-qualitative-only-if-you-cant-count.php">http://www.uxmatters.com/mt/archives/2011/02/usability-testing-is-qualitative-only-if-you-cant-count.php</a>, février 2011.

# UNE DEMARCHE EN QUESTION

S'intéresser aux lacunes des outils permet de déterminer dans quels contextes et dans quelle mesure leur utilisation peut se justifier. L'étude de la démarche des tests A/B s'avère plus fondamentale encore puisqu'elle conduit à se questionner sur leur utilité même. Bien que le marché des tests A/B ait atteint sa maturité outre-Atlantique <sup>84</sup> et que la culture des tests gagne progressivement du terrain en France, il est fondamental de ne pas considérer le recours aux tests A/B comme allant de soi. Au contraire, la popularité croissante de cette approche doit conduire à s'interroger sur les intentions de ses utilisateurs ainsi que sur les risques potentiels qui peuvent en découler.

## Un concept éthiquement critiquable

Les tests A/B peuvent-ils véritablement être considérés comme un outil permettant d'améliorer l'expérience de navigation des utilisateurs d'un site Web ? A qui profite réellement le recours aux tests A/B ? Dans quelle mesure les tests A/B sont-ils encadrés par la loi ? Autant de questions auxquelles il convient de répondre afin de déterminer s'il s'avère pertinent de recourir aux tests A/B, non plus du point de vue des résultats qu'ils permettent d'obtenir mais plutôt du point de vue de l'éthique.

## Des risques de dérives importants

Le recours à un outil de tests A/B, hormis dans un cadre pédagogique, n'est jamais justifié par le souci d'améliorer l'expérience utilisateur. L'expérience utilisateur est systématiquement perçue comme un moyen, un levier permettant d'atteindre un objectif qui la transcende. L'intérêt croissant porté à l'ergonomie et de façon plus générale au design s'explique par la recherche de la performance, dans un contexte de restriction des investissements. De ce fait, l'optimisation de la conversion apparaît comme le moyen le plus simple, le plus rapide et le moins onéreux d'augmenter les volumes de visiteurs convertis tout en maintenant à des niveaux similaires les budgets dédiés à l'acquisition de trafic. Ce constat s'applique tout autant aux organisations à buts lucratifs qu'à celles n'ayant pas fixé la recherche du profit comme justification même à leur existence. Sur le Web tout investissement consenti par une organisation est motivé par la recherche d'un gain. Il n'existe donc pas de particularisme des tests A/B, ni d'angélisme leur étant associés. De ce fait, si les enseignements et optimisations découlant des tests A/B en viennent à profiter aux visiteurs, il ne s'agit pas d'une fin en soi. Partant de ce constat et sachant qu'il s'avère impossible de porter un jugement moral sur une démarche ou un outil en tant que tel, mais qu'il convient plutôt d'évaluer l'usage auguel ils sont employés. que penser des tests A/B?

Les motifs d'expérimentation en ligne sont à bien des égards comparables à ceux justifiant le maintien dans des grandes surfaces de rayons dits traditionnels qui proposent notamment les produits à la découpe et non en libre-service. L'objectif visé est entre autre de disposer d'un espace d'expérimentation permanent et présentant un maximum de difficultés pour juger de la pertinence qu'il y aurait à généraliser certaines nouveautés

<sup>84</sup> https://www.trustradius.com/guides/ab-testing/2014/trends.

à l'ensemble des magasins d'une enseigne 85. Les grandes surfaces ont également recours à des méthodes moins visibles faisant appel à l'inconscient des consommateurs telles que la diffusion de parfums, de fausses promotions, un agencement des locaux savamment réfléchi afin de davantage inciter à l'achat. Ceci explique l'apparition en 1999 du terme « market manipulation » pour désigner ces pratiques. « Cette notion créée par Jon Hanson et Douglas Kysar souligne combien la distribution cherche à exploiter les limites cognitives des consommateurs pour conforter son avantage. Le but est de nous mettre en condition pour être plus disposé à acheter, quitte parfois à organiser des mises en scène ou des opérations trompeuses. » <sup>86</sup> Manipulation : le mot est lâché. Comment en effet ne pas songer à ce terme lorsque les tests A/B promettent de démultiplier les performances d'un design simplement en modifiant la couleur d'un bouton ou son libellé. Que penser d'une modification qui n'a pas touché directement le produit ou le service mis en avant, ni fourni des informations supplémentaires à forte valeur ajoutée au visiteur, mais qui a malgré tout permis d'augmenter le taux de conversion. Dans une approche purement rationnelle, la couleur d'un bouton et son intitulé seraient considérés comme n'ayant aucune influence sur la décision d'un visiteur de souscrire à un produit ou un service. L'expérience démontrant l'inverse, il apparaît que les tests A/B constituent une démarche qui vise sciemment par certaines des optimisations mises en place, à s'adresser à l'inconscient du visiteur. Le fait est que le parallèle entre le « market manipulation » et les tests A/B est saisissant. Cependant, il existe deux distinctions majeures entre les tests A/B et les techniques employées traditionnellement dans le Marketing pour s'adresser à l'inconscient d'un prospect 87:

- Le vecteur de diffusion
- Les méthodes d'évaluation.

La méthode de création traditionnelle d'un contenu Webmarketing repose sur l'identification de profils cibles sur la base de critères sociaux, démographiques et économiques <sup>88</sup>. L'élaboration d'une affiche publicitaire, d'un spot télévisé ou encore l'agencement d'une boutique est rarement l'objet d'un test. Lorsque cela est le cas, un panel représentatif de la cible est interrogé, ce qui aboutit à la collecte de données qualitatives à une faible échelle. La conduite de plusieurs tests est réservée aux campagnes marketing aux budgets les plus élevés. Dans tous les cas, les tests précèdent le lancement du produit, tandis que l'évaluation de l'impact du contenu élaboré s'avère malaisée voire impossible le plus souvent. En effet, il s'avère complexe si ce n'est impossible d'adresser simultanément à une même population dans un magasin, à la télévision, à la radio ou dans la presse plusieurs versions d'un même contenu. Les professionnels du marketing qui officient en dehors de l'écosystème numérique, en particulier le Web, pilotent leurs actions en ayant une visibilité extrêmement limitée sur leur impact. La légitimité de la profession repose avant tout sur des études ayant démontré l'effet du marketing sur l'inconscient, mais avant tout sur la nécessité d'accorder de la

<sup>&</sup>lt;sup>85</sup> Debril Thomas, Dubuission-Quellier Sophie, « « Marée », « charcuterie-traiteur ». Le rayon traditionnel en grande surface, lieu d'expérimentation », *Ethnologie française* 1/2005 (Vol. 35), p. 93-102, <a href="http://www.cairn.info/revue-ethnologie-française-2005-1-page-93.htm">http://www.cairn.info/revue-ethnologie-française-2005-1-page-93.htm</a>.

<sup>&</sup>lt;sup>86</sup> Jean-Marie, Consommateur piégé: les manipulations du merchandising, 3 février 2014, http://www.consoglobe.com/manipulations-du-merchandising-cg.

<sup>&</sup>lt;sup>87</sup> Individu faisant l'objet d'une prospection. Potentiel client, il est le principal objet d'attention des professionnels du Marketing en tant que vecteur principal de croissance d'une organisation.

88 GOWARD CHRIS, *Ibidem*.

visibilité à un produit, un contenu ou un service. Ainsi, hors de l'écosystème numérique, le marketing est employé davantage à des fins de promotion que de conversion. Le grand public est parfaitement conscient que des entreprises, notamment les plus importantes, poursuivent de longue date des investissements dans le seul but de trouver le moyen le plus efficace de les influencer. Cependant, les opinions des prospects potentiels sur cette question sont multiples :

- Un premier groupe, sans se déclarer favorable à ces recherches, comprend le point de vue des entreprises et ne s'y déclare pas hostile.
- Un deuxième groupe désapprouve les démarches visant à manipuler des prospects, mais se déclare à même de s'en prémunir.
- Un troisième groupe juge inacceptable de chercher à manipuler à son insu un individu et se déclare en faveur de l'interdiction pure et simple de ce genre de pratiques.

Le point de vue des deux groupes qui ne souhaitent pas l'interdiction de ces techniques de manipulation découle du manque de visibilité sur l'impact réel des techniques marketing. Dès lors, il est aisé de se croire à l'abri puisqu'informé ou de minorer les effets des optimisations apportées par des professionnels du marketing à un contenu et/ou un dispositif. Il en va tout autrement avec les tests A/B. En dépit des nombreuses lacunes techniques dont souffrent les outils disponibles sur le marché et des nombreuses limitations dans l'interprétation des résultats qui en découlent, les tests A/B permettent d'évaluer à grande échelle un vaste spectre de comportements.

L'absence d'un rejet unanime des tentatives de manipulation s'explique également par la finalité des recherches. Ainsi, la diffusion de musique classique dans des lieux publics notamment les parkings et les gares 90 afin de limiter les comportements agressifs ne soulève aucune opposition. Dans certains contextes, la manipulation à des fins jugées « louables » peut être plus que tolérée : elle peut être appréciée et soutenue. Ce comportement se retrouve dans la sphère numérique. Ainsi, un test A/B a été conduit sur le jeu massivement multi-joueurs en ligne League of Legends sans que les internautes en aient été informés avec pour objectif de diminuer les échanges grossiers et violents sur le chat intégré au jeu. Il est apparu que masquer par défaut le volet dédié au chat conduisait à diminuer de plusieurs dizaines de points de pourcentages la part des sessions de jeu émaillées d'échanges agressifs et injurieux. L'annonce par l'éditeur du jeu du déploiement systématique de cette modification et du test ayant permis d'en valider les effets n'a pas causé de réaction négative. Bien au contraire, l'expérimentation a reçu une large couverture médiatique et a eu bonne presse <sup>91</sup>. En 2014, Facebook conduisait un test A/B dans des conditions similaires, à savoir sans que les utilisateurs testés n'aient été informés d'aucune façon. L'objectif était d'évaluer l'impact de l'inclusion de commentaires positifs, neutres ou négatifs dans les fils de discussion des membres du réseau social sur leur humeur. L'annonce des résultats de l'étude par l'équipe de chercheurs ayant conduit le test, un impact nul pour les notifications neutres ou positives et extrêmement faible pour les notifications à tonalité négative, avait suscité une vague

<sup>&</sup>lt;sup>89</sup> Les commentaires de l'article *Consommateur piégé : les manipulations du merchandising* démontrent l'existence de ces trois courants d'opinion distincts. Cependant, leur identification repose sur les lectures et l'expérience de l'auteur.

<sup>90</sup> BOISSEAU OCEANE, *De la musique pour adoucir les mœurs dans les gares*, 4 juin 2013, <a href="http://www.pratique.fr/actu/musique-adoucir-moeurs-gares-92012.html">http://www.pratique.fr/actu/musique-adoucir-moeurs-gares-92012.html</a>.

<sup>91</sup> BOWMAN NICHOLAS, The Ethics of UX Research, http://www.uxbooth.com/articles/ethics-ux-research/, août 2014.

de protestations et un sentiment majoritaire d'indignation <sup>92</sup>. Facebook n'avait fourni aucune justification à la conduite de cette étude en dehors de la recherche de la connaissance pour la connaissance. Dès lors, le public avait été prompt à envisager des usages potentiellement condamnables et répréhensibles par Facebook des conclusions de l'étude <sup>93</sup>. Comment expliquer que les tests A/B dans leur ensemble n'aient jamais reçu un accueil similaire à celui du test A/B conduit par Facebook, alors même que la raison première de leur utilisation vise à trouver le moyen le plus efficace de convertir un prospect en client/acheteur/consommateur/utilisateur ?

## La transparence : une pratique marginale

Le caractère « confidentiel » des tests A/B joue à n'en pas douter un grand rôle dans ce phénomène. En France, le marché des tests A/B bien qu'en pleine expansion n'est pas encore arrivé à maturité. En effet, à peine 28% des 200 sites Web générant le plus de

trafic en France en 2014 avaient recours à un outil de tests A/B début 2015 <sup>94</sup>. En 2011 encore, dans un article publié sur le Journal du net, l'un des médias en ligne incontournables dans le domaine du Web marketing, le fondateur de la société A/B tasty qui domine le marché français des tests A/B qualifiait cette démarche de « nouvea[u] relais de croissance » <sup>95</sup>. Aux Etats-unis, bien que le marché soit qualifié de mature, les réponses à un questionnaire élaboré par la plateforme de dépôt d'évaluations sur les logiciels Trust Radius fixait à 8,4%

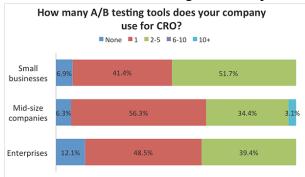


Fig. 5 : Rapport entre le nombre d'outils de tests A/B utilisés et la taille d'une entreprise

Source: *Buyer's guide to A/B testing software* <a href="https://www.trustradius.com/guides/ab-testing/2014/">https://www.trustradius.com/guides/ab-testing/2014/</a>.

le pourcentage de répondants dont l'entreprise n'utilise aucune solution de tests A/B <sup>96</sup>. Si l'utilisation des solutions de tests A/B n'est pas encore un fait acquis en France et aux Etats-Unis pour l'ensemble des professionnels et entreprises présents sur le domaine du Webmarketing, comment espérer que le grand public ait en France une bonne connaissance des tests A/B? Le manque de connaissance du grand public sur le recours aux tests A/B ne serait pas problématique si l'usage des tests A/B en faisait un outil utilisé aux stades préliminaires de l'élaboration d'un site par des professionnels de l'ergonomie et de l'expérience utilisateur. Or, dans les faits il s'agit d'un outil au service de l'optimisation continue utilisé principalement par des professionnels du Web-marketing à des fins d'améliorations de la performance d'un dispositif digital. La participation du

http://www.journaldunet.com/media/publishers/facebook-manipulation-emotionnelle-0614.shtml, juin 2014.

HSIJ JEREMY, Inside the largest virtual psychology lab in the world 27 janyie

<sup>&</sup>lt;sup>92</sup> JAIMES NICOLAS, Facebook est-il coupable de manipulation émotionnelle?,

HSU JEREMY, *Inside the largest virtual psychology lab in the world*, 27 janvier 2015 <a href="https://medium.com/backchannel/inside-the-largest-virtual-psychology-lab-in-the-world-7c0d2c43cda5">https://medium.com/backchannel/inside-the-largest-virtual-psychology-lab-in-the-world-7c0d2c43cda5</a>.

<sup>&</sup>lt;sup>94</sup> JUAN VIRGILE, *Tag manager*, *A/B testing : les parts de marché en France*, 19 janvier 2015, <a href="http://www.journaldunet.com/solutions/analytics/web-analytics-tag-management-a-b-testing-les-parts-de-marche-en-france/part-de-marche-des-outils-de-tag-management-et-d-a-b-testing.shtml">http://www.journaldunet.com/solutions/analytics/web-analytics-tag-management-a-b-testing-les-parts-de-marche-en-france/part-de-marche-des-outils-de-tag-management-et-d-a-b-testing.shtml</a>.

<sup>&</sup>lt;sup>95</sup> AUBERT REMI, *L'A/B testing nouveau relais de croissance des ecommerçants*?, 7 novembre 2011, <a href="http://www.journaldunet.com/ebusiness/expert/50373/l-ab-testing--nouveau-relais-de-croissance-des-ecommercants.shtml">http://www.journaldunet.com/ebusiness/expert/50373/l-ab-testing--nouveau-relais-de-croissance-des-ecommercants.shtml</a>.

<sup>&</sup>lt;sup>96</sup> TRUSTRADIUS, *Buyer's guide to A/B testing software*, <a href="https://www.trustradius.com/guides/ab-testing/2014/">https://www.trustradius.com/guides/ab-testing/2014/</a>, 2014.

grand public est indispensable au bon fonctionnement des tests A/B. Pour l'heure, elle est garantie par l'ignorance du plus grand nombre sur le sujet des tests A/B.

En effet, les entreprises qui choisissent de recourir à ces tests ne cherchent pas à informer leurs visiteurs de leur démarche. Des initiatives d'associations de professionnels des Web analytique ont débouché sur la création de codes éthiques <sup>97</sup>. Néanmoins, les références à ces codes dans les articles et échanges entre les professionnels demeurent marginales. Les éditeurs de sites Web ne les relaient pas sur leur propre plateforme pas plus qu'ils n'incitent leurs employés et prestataires à les respecter. En outre, il n'existe aucun mécanisme contraignant qui pourrait conduire à une hausse de leur adoption. Cet état de fait explique que la transparence soit perçue comme une solution viable à la problématique de la manipulation <sup>98</sup>. En informant le visiteur sur la possibilité de le voir inclus dans un test au cours de sa visite, il deviendrait possible de prévenir tout risque de réaction négative à l'image de celle qu'avait suscité l'expérimentation cachée de Facebook. Dès lors, le visiteur ne subirait plus le test, il en deviendrait un participant consentant et averti. Sans résoudre les problèmes éthiques que soulèvent la finalité du test, l'information du visiteur rendrait la situation en ligne similaire à ce qu'elle peut être hors de l'environnement numérique <sup>99</sup>. Dans le monde réel, l'absence de communication de la part des entreprises sur leur tentative d'influencer le consommateur est contrebalancée par l'ancienneté du phénomène. Connu de longue date, évoqué dans les médias, discuté, il est connu de tous. Il n'en va pas de même en ligne. L'une des qualités des tests A/B qui font défaut aux méthodes d'expérimentation traditionnelles est le caractère passif de la collecte. Le fait que l'individu testé n'ait pas conscience de participer à une expérimentation est l'un des principaux critères concourant à la fiabilisation des résultats. En effet, informer le visiteur n'aurait-il pas pour conséquence de le voir modifier son comportement? La réponse est non car il ne s'agirait pas d'informer les visiteurs du déroulement de chaque test en indiquant son périmètre et les éléments concernés par des modifications. Bien au contraire, les informations à fournir auraient pour objectif d'indiquer que des tests A/B peuvent potentiellement être exécutés sur le site et ce dans un objectif d'amélioration des performances du dispositif. Avec cette approche, l'information du visiteur et la fiabilité des résultats seraient toutes deux garanties. Cependant, faire preuve de transparence n'apporte pas une solution au problème éthique que pose l'utilisation des tests A/B à des fins de manipulation des visiteurs. Qu'en est-il des visiteurs d'un site Web qui rejettent le principe des tests A/B mais qui souhaiteraient malgré tout continuer à utiliser le site?

#### Respecter les directives de la CNIL pour mieux tester ?

Pour véritablement contrebalancer l'absence d'éthique de certaines modifications déployées au moyen de logiciels de tests A/B, il s'avère nécessaire de donner la possibilité à l'utilisateur de pouvoir s'exclure d'un test. A cet effet, les solutions de tests A/B prévoient toutes un mécanisme d'exclusion côté client au moyen d'un *paramètre* 100, ou

<sup>97</sup> DIGITAL ANALYTICS ASSOCIATION, *The Web Analyst's Code of Ethics*, http://www.digitalanalyticsassociation.org/codeofethics, 2011.

<sup>98</sup> BOWMAN NICHOLAS, *Ibidem*.

<sup>99</sup> ESPOSTI SARA DEGLI, *Ibidem*.

<sup>100</sup> Les paramètres servent à transmettre des informations de façon normalisée. Un paramètre est toujours associé à une valeur. Ensemble ils forment un couple où le paramètre constitue une clé donnant accès à une valeur. Les paramètres peuvent être inclus directement dans une URL, auquel cas ils sont insérés après un point d'interrogation, séparés les

d'une fonction Javascript. Cependant, l'implantation d'un dispositif permettant à un internaute de s'exclure d'un test est tellement rare qu'elle peut être qualifiée d'inexistante. Comment expliquer la réticence des entreprises à exploiter une fonctionnalité préembarquée dans l'ensemble des solutions du marché ? L'explication est simple, donner la possibilité aux visiteurs de s'exclure du test risquerait de poser deux problèmes majeurs :

- Un allongement de la durée des tests causée par un manque de visiteurs à inclure dans les variations testées.
- Un risque de manque de représentativité de la population testée. L'un des arguments forts des tests A/B, tester une modification auprès des véritables visiteurs d'un site Web, perdrait de sa pertinence puisqu'une partie de l'audience d'un site risquerait d'être exclue systématiquement des tests. Le problème du silence des non répondants auxquels doivent faire face les professionnels travaillant sur les résultats de sondages en ligne affecterait également les tests A/B.

Faute d'une mise en œuvre spontanée par les entreprises de mécanismes d'exclusions, existe-t-il un cadre législatif venant encadrer la réalisation de tests A/B qui pourrait les y contraindre ? Il n'existe pas en France, en Europe ou aux Etats-Unis de loi dédiée spécifiquement à la réglementation des tests A/B. Cependant, depuis le 1er octobre 2014, la CNIL (Commission Nationale Informatique et Liberté) a commencé à procéder à des contrôles relatifs à l'application par les éditeurs de sites Web de la directive européenne "Paquet Télécom", ainsi que de l'article 32 de la loi informatique et liberté. Ces réglementations ont respectivement été approuvées en 2009, et 2011, pour l'article 32 de la loi informatique et libertés. Il a cependant fallu attendre le mois de Décembre 2013 avant que la CNIL n'en vienne à publier des recommandations techniques et pratiques en vue d'accompagner les éditeurs dans leur mise en conformité avec la loi, premier signe d'une évolution de sa position d'un attentisme bienveillant vers une posture davantage offensive, dont l'annonce de la menée de contrôles effectifs est la manifestation la plus récente. La CNIL dispose d'un argument des plus dissuasifs pour contraindre les éditeurs de sites Web à respecter la législation en vigueur, puisque tout contrevenant à la loi informatique et liberté est susceptible de se voir infliger une amende pouvant atteindre jusqu'à 150 000 euros et ce sans récidive <sup>101</sup>. L'objectif de la loi est triple :

- Encadrer le traçage des internautes par les régies publicitaires et les éditeurs de site Web.
- Contraindre les acteurs ayant recours à la publicité en ligne et/ou utilisant des outils d'analyse comportementale à informer les internautes de la nature des informations collectées.
- Offrir la possibilité aux internautes de ne pas être pistés et tracés au cours de leur navigation.

uns des autres par des « & » et associés à leur valeur au moyen du signe « = ». Ils peuvent également être transmis en « pièce jointe » à une requête, sans être spécifiés dans l'URL.

<sup>101</sup> DEBES FLORIAN, *Cookie la CNIL commence ses contrôles*, 1er octobre 2014, <a href="http://business.lesechos.fr/directions-numeriques/0203817602269-cookies-la-cnil-commence-ses-controles-103790.php">http://business.lesechos.fr/directions-numeriques/0203817602269-cookies-la-cnil-commence-ses-controles-103790.php</a>.

La loi distingue les solutions et dispositifs requérant le recueil du consentement de l'internaute en vue de procéder à une collecte d'information, de celles jugées insuffisamment intrusives pour ne pas justifier le déploiement d'une mesure autre qu'une simple notification à l'internaute. L'utilisation d'un outil de tests A/B suffit à elle seule à faire basculer un site, une application ou un quelconque dispositif y ayant recours dans la première catégorie <sup>102</sup>. De façon plus générale, l'utilisation d'un logiciel de Web analytique qui est devenue systématique sur l'ensemble des sites Web impose à elle seule d'informer le visiteur <sup>103</sup>. Suite à l'avertissement de la CNIL quant au lancement des premiers contrôles, des bandeaux d'information sont apparus sur un nombre croissant de sites. Ils signifient aux visiteurs que des données relatives à leurs habitudes de navigation sont collectées à des fins d'amélioration de l'expérience utilisateur et de ciblage publicitair. En outre, une page d'information dédiée permet aux internautes désireux d'en apprendre davantage ou de ne pas être pistés de s'informer.

Quelles sont les mesures qu'impose la loi et que les éditeurs de sites Web se refusent, pour la plupart, pour l'heure, à déployer ?

- Les solutions proposées à l'internaute afin qu'il soit exclu de la collecte et des tests ne doivent pas l'empêcher de bénéficier des mêmes services que ceux proposés aux internautes ayant accordé leur consentement. Or, la quasi-totalité des sites se limitent à fournir des instructions sur les modalités de blocage des cookies. Non seulement cette approche nécessite des connaissances sur le fonctionnement du Web que n'ont pas la plupart des visiteurs, mais surtout elle s'avère de nature à nuire à l'expérience utilisateur puisque la méthode de blocage des cookies indiquée s'effectue par le truchement des paramètres navigateurs. Or ces derniers permettent seulement de paramétrer des exclusions au niveau d'un nom d'hôte et non d'un cookie en particulier. Sachant que l'authentification sur un site Web repose sur les cookies, pour ne citer que cette fonctionnalité, leur désactivation empêcherait toute implantation d'un espace personnel. Le blocage des cookies en dégradant l'expérience de navigation s'avère incompatible avec la loi informatique et liberté.
- La collecte de données doit être désactivée tant que le consentement de l'internaute n'a pas été donné. La mise en place du message d'avertissement ne s'est pas accompagnée d'un mécanisme de blocage par défaut de la collecte sur la quasitotalité des sites Web sur lesquels il a été implanté. Cette exigence s'avère plus problématique encore que la possibilité donnée à l'utilisateur de choisir à quels outils il accepte de transmettre des données. En effet, les premiers retours indiquent qu'à peine 1% des internautes choisissent de désactiver la collecte de données sur les sites proposant des méthodes simple pour la mettre en œuvre 104. En revanche, avec un taux de rebond moyen de 40% 105 sur le Web, l'application de cette mesure entrainerait des manques irréparables dans la collecte de données à l'occasion d'un test A/B.

105 https://blog.kissmetrics.com/bounce-rate/.

<sup>&</sup>lt;sup>102</sup> A l'exception de l'outil de Web analytique Piwik.

<sup>103</sup> http://www.cnil.fr/vos-obligations/sites-web-cookies-et-autres-traceurs/outils-et-codes-sources/la-mesure-daudience/.

DEBES FLORIAN, Nutella met ses cookies en option, 24 mars 2014, <a href="http://business.lesechos.fr/directions-numeriques/0203336210768-nutella-met-ses-cookies-en-option-61614.php">http://business.lesechos.fr/directions-numeriques/0203336210768-nutella-met-ses-cookies-en-option-61614.php</a>.

• La possibilité de désactiver une solution logicielle tout en conservant une ou plusieurs autres actives. En dépit de la présence de fonctionnalités natives aux solutions de tests A/B la part des sites proposant à l'internaute d'en faire usage est insignifiante.

Le législateur fourni un cadre clair et des solutions techniques concrètes afin de permettre à un utilisateur d'être informé du déroulement de tests A/B sur un site et de pouvoir le cas échéant s'en exclure. Les éditeurs des différentes solutions de tests A/B proposent chacun des méthodes natives à leur outil permettant d'exclure aisément un visiteur d'un test. Dans les faits, la mise en conformité avec la loi est à l'heure actuelle très largement insuffisante sur le Web français. Les propriétaires des sites Web, les agences et les analystes n'ont rien à gagner mais beaucoup à perdre en se mettant en conformité avec la loi. Il semblerait que la mise en conformité ne pourra être réalisée que par la contrainte. Or, pour l'heure, aucune sanction n'a été prononcée par la CNIL bien qu'elle ait annoncé poursuivre ses contrôles <sup>106</sup>.

Certaines modifications évaluées grâce à des outils de tests A/B cherchent à influencer le visiteur au niveau de son inconscient. Les possibilités techniques offertes par le recueil de données quantitatives permettent de quantifier l'impact de ces manipulations de façon plus effective qu'avec les techniques et vecteurs de diffusion traditionnels du Web marketing. Les tests A/B demeurent très largement méconnus du grand public tandis que les éditeurs de sites Web tardent à se mettre en conformité avec la loi qui leur impose pourtant d'informer et de permettre au visiteur de s'exclure d'un test. L'application de la loi permettrait de contrebalancer le manque d'éthique d'une partie des expérimentations réalisées au moyen des tests A/B. La recherche permanente de meilleurs résultats a des conséquences éthiques, mais aussi sur la fiabilité des résultats obtenus suite à la réalisation d'un test A/B.

## Une démarche qui requiert du temps

Les tests A/B offrent la garantie d'adresser à un instant T, à une population de visiteurs constante, des variations d'une page afin d'en comparer les performances avec la version d'origine. Cependant, l'assignation aléatoire d'une variation à un visiteur ne garantit pas nécessairement la constitution d'échantillons de visiteurs similaires d'une variation à une autre. En outre, les tendances observées évoluent fortement au fil du temps. Cette information amène à s'interroger sur l'impact que peut avoir la durée d'un test sur les résultats engrangés. Comment les professionnels et les éditeurs de solutions prennent-ils en compte cet élément lors de la réalisation d'un test ?

#### Le temps, une source majeure d'erreurs d'interprétation

31

La question de l'impact de la durée d'un test sur la validité statistique des résultats engrangés est fondamentale. Des facteurs externes peuvent jouer sur les résultats d'un test et limiter considérablement la portée des analyses réalisables sur les données. Ainsi, si une campagne publicitaire multicanaux est lancée parallèlement au test, les résultats

<sup>&</sup>lt;sup>106</sup> GAVOIS SEBASTIEN, *Bilan 2014 de la CNIL : 5 821 plaintes, 421 contrôles, 18 sanctions et 359 cookies sur une page*, 16 avril 2015, <a href="http://www.nextinpact.com/news/93848-bilan-2014-cnil-5-825-plaintes-421-controles-18-sanctions-et-359-cookies-sur-page.htm">http://www.nextinpact.com/news/93848-bilan-2014-cnil-5-825-plaintes-421-controles-18-sanctions-et-359-cookies-sur-page.htm</a>

seront biaisés. Le site enregistrera un pic de visiteurs dont le comportement et les préoccupations ne reflèteront pas nécessairement celles des visiteurs habituels. En outre, en cas d'achat d'espaces publicitaires sur le Web, leur page d'entrée sur le site risque d'être différente de celle qu'ils auraient empruntée en accédant au site via les résultats naturels 107 d'un moteur de recherche. En effet, lors de la création d'une campagne publicitaire, chaque annonce se voit associer une URL correspondant à la page de destination sur laquelle l'internaute se verra dirigé en cas de clic. Ces pages ne sont pas nécessairement bien classées dans les résultats naturels sur des requêtes ayant déclenché l'affichage des résultats sponsorisés. La prise en compte des marronniers <sup>108</sup>, des effets de saisonnalité et plus généralement des différences de comportements chaque jour de la semaine est également indispensable. Ainsi, les ventes de matériel de ski en France seront toujours plus fortes en hiver qu'en toute autre saison. La consultation d'un site proposant la vente d'équipements en hiver répondra davantage à un impératif d'achat immédiat tandis qu'au printemps, en été et en automne, la nécessité d'acheter à court terme sera moins forte. La visite répondra davantage à une problématique de recherche d'informations, et/ou de promotions. La mise en ligne d'une version responsive ou au contraire son absence dans un contexte d'augmentation des visiteurs mobiles peut également conduire à altérer significativement les performances enregistrées. Enfin, l'introduction d'une modification peut avoir à court terme un impact positif ou à l'inverse négatif sur les visiteurs récurrents du site. Néanmoins, l'effet de la nouveauté s'estompant avec le temps, les résultats enregistrés par une variation pourront significativement évoluer avec le temps. Le temps peut dont remettre en question la fiabilité des données collectées lors d'un test :

- En faussant les résultats engrangés durant le test à la hausse ou à la baisse du fait de facteur externes.
- En donnant des tendances à la hausse ou à la baisse d'une variation du fait d'une durée de test insuffisante pour permettre de lisser les effets induits par la nouveauté.

Plusieurs solutions permettent de lutter contre les facteurs perturbateurs externes. L'exclusion des internautes en provenance de certaines campagnes publicitaires permet d'éviter qu'ils pèsent sur les résultats engrangés. L'exclusion pure et simple de visiteurs acquis grâce à des annonces publicitaires peut également être une solution viable sur des sites Web ne réalisant qu'occasionnellement des achats d'espaces en ligne. En revanche, il n'existe pas de solution simple à mettre en œuvre afin de lutter contre les marronniers et les comportements spécifiques à un jour de la semaine ou une période de l'année. A défaut de pouvoir pondérer les résultats au moyen d'une formule mathématique, l'analyste doit avoir une parfaite connaissance du secteur d'activité et de la cible du site Web sur lequel s'effectue le test. Si l'analyste chargé du test appartient à l'organisation propriétaire du site Web, il lui sera plus facile de tenir compte des éléments externes propres au secteur d'activité de son organisation. En revanche, si la réalisation des tests est sous-traitée, intégralement ou en partie, le risque d'erreur sera beaucoup plus élevé. La taille, la composition et l'expérience de l'équipe chargée des tests A/B apparaissent

\_

<sup>107</sup> Les résultats naturels sont les liens vers des contenus retournés par un moteur de recherche dans une page de résultats en réponse à une requête. Le classement et la sélection des résultats sont déterminés par l'algorithme du moteur de recherche. Aucun investissement n'est nécessaire pour figurer dans cette liste de résultats. En revanche, les résultats sponsorisés sont des emplacements publicitaires dont l'attribution est gérée par un système d'enchères. L'annonceur au budget le plus élevé occupera l'emplacement le plus élevé au sein de la zone réservée aux résultats sponsorisés.

<sup>&</sup>lt;sup>108</sup> Le terme marronnier désigne des phénomènes récurrents se reproduisant toujours à une date ou sur une plage de date identique.

comme des facteurs décisifs pour aboutir à une bonne interprétation des données. Le risque d'erreur est inversement proportionnel au nombre d'intervenants sur un projet. Le manque d'expérience et/ou de connaissances peut aboutir à des résultats désastreux tels que la mise en ligne d'une variation en apparence gagnante mais qui s'inscrit en réalité en régression par rapport à l'original. Ce genre de scénario est amené à se produire invariablement dans les cas de figure suivants :

- Séduits par le discours des éditeurs de solutions <sup>109</sup> quant à la facilité de prise en main et les possibilités d'usage pour des profils non-techniques, des spécialistes du marketing rattachés à l'équipe responsable d'un site Web peuvent décider de se charger eux même des tests. Dans le cadre de la souscription d'un abonnement entreprise, les conseils d'un consultant de l'éditeur de la solution de tests A/B pourraient leur éviter de commettre certaines erreurs. Cependant, en l'absence d'accompagnement, en supposant qu'aucune erreur n'ait été commise lors de la création et du paramétrage du test <sup>110</sup>, faute d'expérience et de connaissances suffisantes sur les limites du logiciel utilisé des erreurs seront immanquablement commises lors de l'analyse. Dans ce scénario, la facilité de création des tests et le discours rassurant des éditeurs de solutions conduisent à sous-estimer la complexité d'une démarche de tests A/B.
- Les responsables d'un site Web peuvent décider de faire appel à une équipe d'experts des tests A/B et ou à un analyste indépendant. Les personnels en charge de la création, du paramétrage et de l'analyse des tests impliquent les équipes du site Web uniquement aux stades de la définition des objectifs et de la validation des designs définitifs des variations. Les échanges ayant été limités, les analystes n'ont pu acquérir une connaissance suffisante des spécificités du secteur d'activité de leur employeur. Confiants dans l'expertise du ou des analystes, les responsables du site Web ne consultent pas les données brutes du test. Dès lors, ils s'avèrent le plus souvent dans l'impossibilité de détecter des signaux anormaux lorsque des données remises en forme et commentées leur sont présentées. Dans ce scénario, la croyance en l'infaillibilité des analystes et un manque d'implication des responsables du site conduiront à des erreurs d'interprétation.

Cependant, plusieurs modes d'organisation permettent de limiter considérablement les risques d'erreurs d'interprétation induits par des facteurs externes :

• L'équipe chargée de la mise à jour et de l'actualisation du site Web est sensible aux questions de l'amélioration continue et des tests. Elle comprend au minimum un spécialiste de la conversion au fait des spécificités du secteur d'activité de l'entreprise.

Pour A/B Tasty voir: <a href="https://fr.abtasty.com/features#feature">https://fr.abtasty.com/features#feature</a> editor.

<sup>109</sup> A ce stade du mémoire, le lecteur pourra être étonné de trouver une note renvoyant vers un thème déjà abordé à de multiples reprises. Cette fourniture tardive de la référence s'explique par la nécessité d'avoir été sensibilisé préalablement aux limites des tests A/B. Avoir connaissance du point de vue du parti opposé est une condition indispensable à la réalisation d'une lecture critique d'un document. Deux pages d'informations généralistes des deux solutions leaders en France donnent une idée assez précise du discours enthousiaste et peu nuancé des revendeurs de solutions :

Pour Optimizely voir : <a href="https://www.optimizely.fr/benefits/">https://www.optimizely.fr/benefits/</a>.

<sup>&</sup>lt;sup>110</sup> Optimizely est l'acteur proposant la base de connaissance la plus détaillée et la plus fournie sur son logiciel et la démarche de tests A/B en général. « L'académie » d'Optimizely est un ensemble d'articles et d'activités conçus dans l'optique d'accompagner un néophyte dans sa prise en main de la solution de l'éditeur. Les contenus proposés ont été élaborés en sélectionnant dans la base de connaissance de l'éditeur, beaucoup plus vaste, les éléments les plus indispensables. Difficile d'imaginer lancer un test en quelques minutes lorsque le logiciel permettant de les créer et d'en contrôler les effets nécessite de prendre connaissance au minimum de quatre modules pour pouvoir être utilisé. Voir : https://learn.optimizely.com/hc/en-us.

• L'équipe en charge du site Web n'appartient pas à une organisation mûre du point de vue des processus d'optimisation en continu. Cependant, la volonté de changement est forte en interne. L'appel à des compétences externes pour la réalisation de tests s'accompagne d'un dialogue constant entre les deux parties qui sont toutes deux impliquées à chacune des phases du projet.

S'il est possible de lutter efficacement contre les mauvaises analyses induites par des facteurs externes conjoncturels ou propres à un secteur d'activité en multipliant les intervenants et les échanges sur un projet, gérer le phénomène de la nouveauté s'avère bien plus complexe. Quelles réponses existe-t-il à ce problème ?

#### Les calculateurs d'échantillon : une réponse unique mais insatisfaisante

A chaque réalisation d'un test A/B, le même phénomène se reproduit invariablement : les tendances mesurées quotidiennement se reproduisent rarement à l'identique le ou les jours suivant(s). Les résultats d'un test sont une moyenne des comportements journaliers. Durant les premiers jours d'un test, il est fréquent de constater

des inversions de tendances multiples, suivies par une stabilisation temporaire des enregistrés, résultats d'un accroissement d'une tendance à la hausse/baisse. Les performances d'une variation ne sont pas constantes dans le temps. Le graphique de la figure 3 illustre le problème auquel



Fig. 6 : Courbe d'évolution du taux de conversion quotidien d'un objectif

Source: Interface de rapports de la solution A/B Tasty

est confrontée toute organisation ayant recours aux tests A/B. Faut-il considérer comme acquise la tendance observée durant quatre jours à compter du 16 avril et qui tranche nettement avec les fortes fluctuations des cinq premiers jours par sa constance ? Quelle(s) réponse(s) les professionnels des tests A/B ont-ils apporté à la question de la durée idéale d'un test pour éliminer tout risque d'erreurs induit par l'introduction de nouveaux éléments et l'impératif de validité statistique des résultats ? La réponse est simple : il n'v en a pas. Il n'existe pas de consensus autour de la question de la durée minimale d'un test. Ainsi, il est révélateur de constater que le contenu de la page Web hébergée sur le site d'Optimizely dont le titre est « How long to run a test » <sup>111</sup> ne parle à aucun moment de la durée d'un test en tant que telle. L'approche d'Optimizely est la même que celle de l'ensemble des éditeurs de solutions de tests A/B et peut être résumée de la façon suivante : la valeur des données collectées à l'occasion des tests A/B repose sur leur évaluation par un algorithme qui permet de déterminer leur fiabilité d'un point de vue statistique. Chaque test étant différent, il est impossible de fournir une règle efficace de calcul de la durée d'un test. En revanche, il est possible de déterminer le volume minimum de visiteurs à affecter à la version d'origine, ainsi qu'aux variations pour garantir la fiabilité statistique des résultats.

L'évaluation de l'échantillon minimum de visiteurs requis est possible sans connaissance techniques particulières par l'intermédiaire d'un calculateur que mettent à disposition

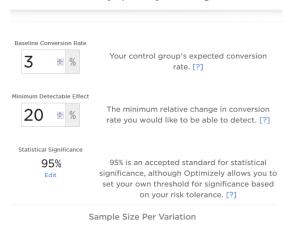
https://help.optimizely.com/hc/en-us/articles/200133789-How-long-to-run-a-test.

tous les éditeurs. Pour pouvoir fournir une estimation, il faut être en mesure de connaître le taux de conversion d'origine de l'indicateur qui servira de référence pour déterminer si un test s'avère être un succès ou un échec. Ensuite, il convient d'indiquer le pourcentage d'évolution minimum à partir duquel une évolution des performances pourra être jugée satisfaisante. Enfin, indiquer la pertinence statistique désirée <sup>112</sup>. Les limites de cette approche sont multiples :

- Les problèmes liés à la durée d'une expérimentation en dehors de l'aspect statistique ne sont pas pris en compte.
- L'utilisation de l'outil nécessite de connaître le taux de conversion d'origine de l'indicateur. Autrement dit de disposer d'un outil de Web

#### A/B Test Sample Size Calculator

Powered by Optimizely's Stats Engine



10.170

Fig. 7 : Le calculateur d'échantillon d'Optimizely Source : Interface de rapports de la solution A/B Tasty

Analytique sur son site et d'un historique de données conséquent. Cependant, certains indicateurs non suivis par défaut dans les outils de Web Analytique peuvent empêcher de disposer de ces données. Exemple : si un test avait pour objectif principal l'augmentation du taux de clics sur des accordéons servant à masquer des contenus textuels. En l'absence d'un plan de marquage avancé, cette information ne serait pas disponible dans l'outil de Web Analytique. Il s'avèrerait donc nécessaire de modifier le plan de marquage du site, puis d'attendre plusieurs mois avant de disposer d'une moyenne suffisamment représentative.

• Au-delà des problèmes de calcul du taux de conversion d'origine des indicateurs de performance principaux du test, il s'avère nécessaire d'adopter la même démarche pour chacun des indicateurs secondaires que l'analyste sera amené à examiner afin de déterminer si le test n'a pas eu des conséquences imprévues. Autrement dit, le calcul de l'échantillon minimum n'est pas nécessairement déterminé par les objectifs principaux du test, mais plutôt par l'indicateur nécessitant l'assignation du nombre le plus important de visiteurs à une variation. Or, l'intérêt pour ces indicateurs secondaires apparaît le plus souvent au stade de l'analyse. Il s'agit le plus souvent d'objectifs non-rétroactifs. Par conséquent, il s'avèrera nécessaire de se tourner vers l'outil de Web Analytique pour avoir de la visibilité sur leur évolution, mais sans pouvoir évaluer la fiabilité statistique des résultats enregistrés. Un objectif rétroactif pourrait potentiellement poser un problème identique dans la mesure où il nécessiterait de relancer le test afin d'atteindre un volume supérieur de visiteurs minimums requis.

\_\_\_

<sup>112</sup> Pour en savoir plus sur chacun de ces éléments voir : <a href="https://help.optimizely.com/hc/en-us/articles/200133789-How-long-to-run-a-test#minimum">https://help.optimizely.com/hc/en-us/articles/200133789-How-long-to-run-a-test#minimum</a>. A noter, depuis le mois de janvier 2015 Optimizely est le seul acteur du marché à proposer un algorithme capable de déterminer automatiquement à partir de quel volume de visiteurs les résultats peuvent être déclarés valides d'un point de vue statistique. Désormais, le calculateur ne sert plus à déterminer la date d'arrêt d'un test, mais uniquement à prévoir sa durée. Pour en savoir plus voir : <a href="http://blog.optimizely.com/2015/01/20/statistics-for-the-internet-age-the-story-behind-optimizelys-new-stats-engine/">https://blog.optimizely.com/2015/01/20/statistics-for-the-internet-age-the-story-behind-optimizelys-new-stats-engine/</a>.

Afin de s'assurer que les effets de la nouveauté, ou de facteurs externes non pris en compte telle que l'initiative d'un concurrent, soient les plus faibles possibles l'allongement de la durée des tests apparaît comme la seule solution viable et efficace.

#### Allonger les durées pour mieux tester ?

Bien que les calculateurs ne soient pas exempts de défauts, ils permettent de déterminer une durée minimum en dessous de laquelle il convient de ne tomber sous aucun prétexte. Cependant, une fois l'échantillon minimum de visiteurs atteint, le test doit entrer dans une seconde phase dite d'allongement. Désormais, l'amélioration ou la dégradation des indicateurs de performance signifiera la remise en question de la validité du test. Quant aux variations minimes ou mieux encore au maintien des performances, ils seront le signe que les premiers résultats enregistrés étaient les bons. Cependant, il s'avère impossible de déterminer combien de temps doit durer cette seconde phase. Certains analystes recommandent, pour contrôler la validité d'un test, d'en lancer un second, identique en tous points au premier, une fois l'échantillon minimum de visiteurs atteint. Cependant, cette approche ne saurait être retenue comme une réponse viable au problème dans la mesure où le second test sera sujet aux mêmes limitations que le premier. L'expérience et l'empirisme apparaissent comme les deux seules solutions partielles à ce problème. Eux seuls permettront à l'analyste de décider du moment opportun pour mettre fin à la seconde phase de l'expérimentation, dans l'hypothèse où les courbes de performances seraient restées constantes. Si cette approche permet de limiter les risques d'erreurs et de diminuer la part des faux positifs, peut-elle pour autant être appliquée en situation réelle?

L'allongement de la durée des tests ferait perdre aux tests A/B leur statut d'outil agile au service de processus d'amélioration continue reposant sur des cycles itératifs courts <sup>113</sup>. En outre, elle s'avèrerait totalement incompatible avec les problématiques des principaux utilisateurs des tests A/B : les professionnels du Webmarketing <sup>114</sup>. Pour rappel, ces derniers voient dans le recours aux tests A/B un outil à même de les aider à atteindre leurs objectifs d'amélioration de la performance. Or l'allongement des tests, du fait des biais qu'introduirait la conduite simultanée de plusieurs tests, limiterait considérablement le nombre de tests réalisables en une année. En outre, en cas de détection d'un faux positif grâce à l'allongement de la durée d'un test, le responsable des tests se trouve confronté à une situation doublement problématique :

- L'échec du test n'ayant pas remis en question son impératif de résultats, il devra malgré tout y parvenir après avoir perdu un temps précieux. Si la démarche de test n'est pas majoritairement soutenue au sein de l'entreprise, la pression pour obtenir des résultats s'en trouvera accrue.
- Dans le compte rendu du test, l'analyste devra expliquer comment la situation a pu évoluer entre des premiers rapports dépeignant une tendance positive et des résultats finaux négatifs. Ce faisant, il risque d'instaurer un doute quant à la fiabilité même de la démarche.

\_

<sup>&</sup>lt;sup>113</sup> Les tests A/B sont utilisés avant tout pour la mise en place de modifications mineures à intervalles réguliers, les gains dépassant largement l'investissement requis pour leur implémentation. Néanmoins, à moyen/long terme cette approche n'est pas nécessairement la plus susceptible d'offrir le meilleur retour sur investissement. Pour en savoir plus, voir : PORTER JOSHUA, *The Local Maximum*, <a href="http://52weeksofux.com/post/694598769/the-local-maximum">http://52weeksofux.com/post/694598769/the-local-maximum</a>, juin 2010

<sup>114</sup> KURT WILL, *Ibidem*.

L'allongement de la durée des tests A/B dans un contexte de recherche d'un retour sur investissements rapide ou plus simplement de pressions fortes autour de l'obtention résultats apparaît inadapté. Il existe une opposition intrinsèque entre la méthode opératoire permettant de fiabiliser les résultats d'un test A/B et les besoins auxquels leur utilisation permet de répondre.

Une prise de conscience des annonceurs quant à la difficulté de déterminer le bon moment pour mettre fin à un test est peu probable. En effet, elle nécessiterait un travail de sensibilisation des éditeurs de solutions d'une part, et des professionnels du Webmarketing d'autre part qui auraient tous deux beaucoup à perdre et rien à gagner dans cette démarche. A court terme, en l'absence de visibilité sur une percée technologique ou méthodologique majeure à même de résoudre ce problème, il semblerait qu'il soit amené à perdurer. Des résultats dont l'évaluation du niveau de fiabilité demeurera imparfaite continueront à être présentés comme indiscutables en vertu de leur conformité à des méthodes d'évaluation statistiques. La culture de confiance dans la fiabilité des données quantitatives et dans les méthodes automatisées de vérification qui leur sont adossées conduit ainsi à adopter une approche reposant davantage sur la croyance que la raison.

D'une part, les tests A/B peuvent servir à tester le déploiement de modifications visant à manipuler un visiteur, ce qui s'avère éthiquement condamnable. D'autre part, la fiabilité des données collectées s'avère fortement sujette à des facteurs externes, notamment le temps. Dans un cas comme dans l'autre, le meilleur moyen de lutter contre ces deux phénomènes est l'adoption d'une démarche proactive sous la forme de mesures concrètes à prendre dans le cadre de la conduite de tests A/B. Cependant, l'instauration de ces mesures étant conditionnée à des actions concrètes de la part des responsables de sites Web et/ou des analystes, qui n'ont aucun intérêt à les prendre, elles demeurent inappliquées dans l'écrasante majorité des cas. Conséquence de cet état de fait, le recours aux tests A/B apparaît à l'heure actuelle non conforme aux règles de l'éthique professionnelle et de la morale.

# **CONCLUSION**

Les tests A/B sont un outil à la mode destiné à répondre à des problématiques à la frontière du Webmarketing, de l'expérience utilisateur, de l'optimisation de la conversion et de l'architecture de l'information. La croissance soutenue de ce marché, qui est parvenu à maturité outre-Atlantique tandis qu'il demeure en maturation en France, a conduit à une explosion de l'offre. Plusieurs dizaines de solutions dédiées aux tests A/B existent sur le marché tandis qu'un nombre croissant de produits logiciels ajoutent une brique de test à leurs fonctionnalités natives. Au-delà de l'effet de mode, l'engouement pour les tests A/B repose sur une promesse de retour sur investissement rapide basée sur un déploiement et une utilisation ne nécessitant aucune compétence technique particulière. Surtout, les tests A/B sont valorisés et présentés comme une approche novatrice en rupture avec les méthodes traditionnelles dédiées à la conception et l'optimisation d'interfaces utilisateurs du fait d'un usage intensif de la donnée. Grâce à leurs algorithmes évaluant les actions des utilisateurs sur un site Web ou une application, les solutions de tests A/B seraient en mesure de rendre un jugement objectif et incontestable dans le but de sélectionner le meilleur design possible dans le cadre d'une mise en compétition de versions alternatives d'une page par rapport à la version d'origine. Grâce à la collecte de données quantitatives, évaluées au moyen de méthodes statistiques constantes et clairement formalisées, les tests A/B permettraient le passage d'une conception basée sur l'intuition, le ressenti et l'expérience à celle des données, des faits et des statistiques. Surtout, la quantification des progressions enregistrées permettrait de rendre compte plus justement des apports des modifications affectant l'expérience utilisateur et de l'extrême rentabilité des méthodes d'optimisation de la conversion reposant sur l'usage des tests A/B.

Or, il apparaît après examen qu'il existe un décalage entre la portée et la valeur réelle des données collectées à l'occasion d'un test A/B et ce qu'elles sont censées être et permettre d'apprendre. Ainsi, les tests A/B ne rendent pas compte des comportements d'un utilisateur, mais plutôt du ou des utilisateurs d'un navigateur. Sachant que la collecte d'informations et l'identification des navigateurs reposent exclusivement sur des cookies dont la suppression est courante, que plusieurs utilisateurs peuvent utiliser un même navigateur et plusieurs appareils être utilisés par un même individu, il apparaît qu'il existe un grand décalage entre l'acceptation courante du terme individu et le sens que ce terme recouvre dans le cadre d'un test A/B. A cette limitation technique d'envergure s'en ajoutent d'autres liées aux méthodes de collecte de l'information qui peuvent rendre impossible l'enregistrement de certains types d'actions utilisateurs. En outre, la collecte exclusive de données quantitatives par les outils de tests A/B limite considérablement la valeur et la portée des analyses pouvant en découler quand elle ne les fausse pas purement et simplement. A ces problèmes d'outils s'en ajoutent d'autres liés à la conduite des tests proprement dite. En effet, les modifications évaluées grâce aux tests A/B s'adressent le plus souvent à l'inconscient du visiteur. Ainsi, l'utilisation des tests A/B à des fins de manipulation des utilisateurs d'un site est une démarche banale. Or, contrairement aux méthodes du marketing traditionnel, les internautes ne sont qu'une minorité à avoir conscience de l'existence de ces tests. Les professionnels ayant fréquemment recours aux tests A/B sont le plus souvent soumis à des impératifs de performances des plus exigeants. De ce fait, ils sont contraints de restreindre la durée d'exécution d'un test afin de pouvoir atteindre leurs objectifs, au risque de tirer des conclusions erronées.

Les limites des tests A/B sont donc multiples et trouvent leur origine tant dans les lacunes techniques des outils permettant leur exécution, que dans les nombreux problèmes soulevés par la conduite d'une démarche de tests. Les conséquences de ces

problèmes sont variables. Elles peuvent aller de la diminution des possibilités offertes par les tests A/B dans le cas des limitations techniques des outils, à une remise en question de la véracité des résultats enregistrés lors d'un test A/B dans le cas de la durée des tests, voire à un abandon pur et simple des tests avec les problèmes éthiques soulevés par les manipulations émotionnelles que permettent les tests A/B. Pour l'heure, il n'existe pas de solution pleinement satisfaisante à aucun de ces problèmes, mais plutôt des palliatifs plus ou moins efficaces. Dans le cas des problèmes éthiques soulevés par les tests, l'application de la réglementation en vigueur prévoyant une désactivation par défaut de la collecte de données et une information des visiteurs permettrait de rendre du pouvoir aux internautes sans remettre en question le caractère condamnable de certaines expérimentations. En ce qui concerne les limites des données quantitatives, l'adjonction de données qualitatives via des entretiens utilisateurs ou la fusion d'ensemble de données devraient permettre de limiter les erreurs d'analyse et de mieux comprendre les résultats d'un test. Quant aux limitations techniques des outils, elles semblent impossibles à surmonter à court terme autrement que par une restriction du champ d'étude aux seuls visiteurs documentés par plusieurs sources de données. A moyen terme, des innovations technologiques devraient permettre de corriger certains problèmes, notamment celui de l'authentification. Enfin, le problème de la durée juste d'un test semble le plus difficile à surmonter car il impliquerait une remise en question de certains acteurs, un travail d'évangélisation et une augmentation significatives des coûts qui seraient tous préjudiciables aux tests A/B.

Ces différents constats, bien certains soient que assez préoccupants ne doivent pas conduire à rejeter les tests A/B. Comme l'ensemble des méthodes dédiées à l'étude de l'expérience utilisateur, les tests A/B comportent des défauts dont il convient d'avoir connaissance avant d'y avoir recours afin d'éviter désillusions. erreurs et  $\Pi$ surprenant de constater qu'en dépit de près d'une décennie d'expansion, de l'existence de bibliothèques de cas d'usages 115, de prises de paroles d'experts démontrant les limites des tests A/B, le crédit dont bénéficient restent toujours aussi considérable. Cependant, à mesure que le nombre d'utilisateurs s'accroit

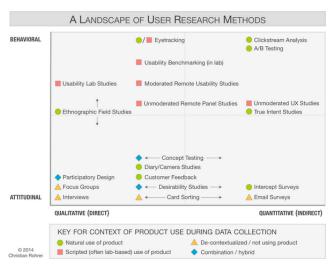


Fig. 8 : Finalité des méthodes d'étude de l'expérience utilisateur

Source : When to Use Which User-Experience Research Methods

http://www.nngroup.com/articles/which-ux-research-methods/

la probabilité d'une prise de conscience du plus grand nombre et d'un regain du discours critique à l'encontre des tests A/B augmente. La pérennité de l'usage des tests A/B repose sur leur démystification.

 $<sup>{\</sup>color{red}^{115}}\ \underline{\text{http://yourstory.com/2014/09/vwo-concept-feedback-acquisition/.}}$ 

Dans ce contexte, les professionnels de l'architecture de l'information peuvent jouer un rôle déterminant. Ils doivent continuer à endosser le rôle de « bâtisseur de ponts » <sup>116</sup> qui leur est traditionnellement dévolu :

- Entre les outils/contenus et les utilisateurs, en menant un travail de démystification et d'évangélisation sur les limites et les atouts réels des tests A/B.
- Entre la recherche et la pratique en partageant leurs méthodes et leurs meilleures pratiques avec le plus grand nombre afin de faire progresser la discipline.
- Entre les plateformes et les canaux en réconciliant les différents outils et les sources de données par une analyse prudente.
- Entre les stratégies et les tactiques en recommandant l'usage des tests A/B uniquement dans les contextes où leur usage s'avèrera pertinent.
- Entre les individus et les disciplines en impliquant les différents acteurs d'un projet dans la démarche de tests A/B.

Les tests A/B constituent une opportunité pour les architectes de l'information dans la mesure où leurs différentes connaissances, compétences et savoir-faire recoupent précisément ceux qui s'avèrent nécessaires pour tirer le meilleur parti des tests A/B.

\_\_\_

<sup>&</sup>lt;sup>116</sup> MORVILLE PETER, *Understanding information architecture*, <a href="https://prezi.com/aafmvya6bk7t/understanding-information-architecture/">https://prezi.com/aafmvya6bk7t/understanding-information-architecture/</a>, août 2014.

# **BIBLIOGRAPHIE**

# > SUR LES ETUDES QUANTITATIVES

## **GENERALITES**

ACQUIA, *Examples of Big Data Projects*, <a href="http://www.acquia.com/examples-big-data-projects">http://www.acquia.com/examples-big-data-projects</a>, 2014.

INFODESIGN, *Putting people first Philips Design: From data to meaning for people*, <a href="http://www.experientia.com/blog/philips-design-from-data-to-meaning-for-people/">http://www.experientia.com/blog/philips-design-from-data-to-meaning-for-people/</a>, 4 novembre 2012.

LIN JIMMY, WONGSUPHASAWAT KRIST, Using Visualizations to Monitor Changes and Harvest Insights from a Global-Scale Logging Infrastructure at Twitter,

http://www.umiacs.umd.edu/~jimmylin/publications/Wongsuphasawat\_Lin\_VAST2014.pdf, 13 novembre 2014.

PAVLISCAK PAMELA, Data-Informed Design: Three Data Stories,

http://www.uxmatters.com/mt/archives/2014/11/data-informed-design-three-data-stories.php, 3 novembre 2014.

TRUSTRADIUS, Buyer's guide to A/B testing software,

https://www.trustradius.com/guides/ab-testing/2014/, 2014.

## PROBLÈMES ETHIQUES

BARANIUK CHRIS, *Internet: How websites are experimenting on you*, <a href="http://www.bbc.com/future/story/20140801-hidden-tricks-that-make-you-click">http://www.bbc.com/future/story/20140801-hidden-tricks-that-make-you-click</a>, août 2014.

BOWMAN NICHOLAS, The Ethics of UX Research,

http://www.uxbooth.com/articles/ethics-ux-research/, août 2014.

DARK PATTERNS, Journey Map, http://darkpatterns.org/journey\_map/, 2014.

DEBES FLORIAN, « Nutella met ses cookies en option » in : Les Echos,

http://business.lesechos.fr/directions-numeriques/0203336210768-nutella-met-ses-cookies-en-option-61614.php, 24 mars 2014.

DIGITAL ANALYTICS ASSOCIATION, The Web Analyst's Code of Ethics,

http://www.digitalanalyticsassociation.org/codeofethics, 2011.

ESPOSTI SARA DEGLI, « When big data meets dataveillance: the hidden side of analytics », in: *Surveillance & Society*, vol. 12, n° 2, mai 2014, pp. 209-225.

JAIMES NICOLAS, Facebook est-il coupable de manipulation émotionnelle?,

http://www.journaldunet.com/media/publishers/facebook-manipulation-emotionnelle-0614.shtml, juin 2014.

# > SUR LES LIMITES DES ETUDES QUANTITATIVES

# EN GÉNÉRAL

BROOKS DATA, « What You'll Do Next », in: The New York Times,

http://www.nytimes.com/2013/04/16/opinion/brooks-what-youll-do-next.html, avril 2013.

CRAWFORD KATE, *The Hidden Biases in Big Data*, <a href="https://hbr.org/2013/04/the-hidden-biases-in-big-data">https://hbr.org/2013/04/the-hidden-biases-in-big-data</a>, avril 2013.

DIRESTA RENEE, *Big data and the « Big Lie »: the challenges facing big brand marketers*, <a href="http://radar.oreilly.com/2013/09/big-data-and-the-big-lie-the-challenges-facing-big-brand-marketers.html">http://radar.oreilly.com/2013/09/big-data-and-the-big-lie-the-challenges-facing-big-brand-marketers.html</a>, 25 septembre 2013.

FRYE PATRICK, *Human Face Of Big Data: Crowdsourcing The World*, http://www.inquisitr.com/352203/human-face-of-big-data-crowdsourcing-the-world/, 3

octobre 2012.

HARDY QUENTIN, Why Big Data Is Not Truth,

http://bits.blogs.nytimes.com/2013/06/01/why-big-data-is-not-truth/, juin 2013.

HARRIS DERICK, *Liking curly fries might not mean you're smart : When mere data isn't enough*, <a href="https://gigaom.com/2013/03/25/liking-curly-fries-might-not-mean-youre-smart-when-correlation-isnt-enough">https://gigaom.com/2013/03/25/liking-curly-fries-might-not-mean-youre-smart-when-correlation-isnt-enough</a>, 25 mars 2013.

MINKEL ELIZABETH, *On Nate Silver's FiveThirtyEight: is it possible for a data-driven journalist to tell a good story*, <a href="http://www.newstatesman.com/2014/03/nate-silver-and-fivethirtyeight-it-possible-data-driven-journalist-tell-good-story">http://www.newstatesman.com/2014/03/nate-silver-and-fivethirtyeight-it-possible-data-driven-journalist-tell-good-story</a>, 25 mars 2014.

MOUTON ANDRE, Does big data have us « fooled by randomness »?,

http://www.usatoday.com/story/tech/2013/08/19/big-data-random-facebook-nsa-linkedin/2670667/, août 2013.

NIELSEN JAKOB, *Risks of Quantitative Studies*, <a href="http://www.nngroup.com/articles/risks-of-quantitative-studies">http://www.nngroup.com/articles/risks-of-quantitative-studies</a>, 1 mars 2004.

SELMAN BILL, Why Do We Conduct Qualitative User Research?,

https://blog.mozilla.org/ux/2014/10/why-do-we-conduct-qualitative-user-research/, 30 octobre 2014.

WANG TRICIA, The Conceit of Oracles: How we ended of up in a world where quantitative data is more valued than qualitative data,

http://triciawang.com/updates/2014/2/5/transcript-of-my-talk-the-conceit-of-oracles-how-we-ended-of.html, février 2014.

## DES TESTS A/B

GATTIKER E. URS, Why A/B testing of web design fails,

<u>http://commetrics.com/articles/supplement-with-qualitative-assessment/</u>, décembre 2009.

KAISER FUNG, The reality is most A/B tests fail, and Facebook is here to help,

http://junkcharts.typepad.com/numbersruleyourworld/2014/04/the-reality-is-most-abtests-fail-and-facebook-is-here-to-help.html, avril 2014.

KURT WILL, Most of Your A/B Test Results Are Illusory and That's Okay.

https://blog.kissmetrics.com/your-ab-tests-are-illusory/, avril 2014.

NIELSEN JAKOB, Putting A/B Testing in Its Place,

http://www.nngroup.com/articles/putting-ab-testing-in-its-place/, août 2005.

## > SUR LES MOYENS DE DEPASSER CES LIMITES

### ASSOCIER DONNEES QUANTITATIVES ET QUALITATIVES

ALHADI FARID, Combining Qualitative and Quantitative Data to Gain Useful Insights, <a href="http://www.e-nor.com/blog/web-analytics/combining-qualitative-and-quantitative-data-to-gain-useful-insights">http://www.e-nor.com/blog/web-analytics/combining-qualitative-and-quantitative-data-to-gain-useful-insights</a>, 8 mars 2013.

BERKELEY UNIVERSITY, *Ethno-Mining*: *Integrating Numbers and Words from the Ground Up*, <a href="http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-125.pdf">http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-125.pdf</a>, 6 octobre 2006.

DRISCOLL DAVID L.; APPIAH-YEBOAH AFUA; SALIB PHILIP; RUPERT DOUGLAS J.,

« Merging Qualitative and Quantitative Data », in : *Mixed Methods Research : How To and Why Not*, University of Georgia, vol. 18, 2007,

http://digitalcommons.unl.edu/icwdmeea/18.

GIRARDIN FABIEN, *Insights from network data analysis that yield field observations*, <a href="http://ethnographymatters.net/blog/2013/04/02/insights-from-network-data-analysis-that-yield-field-observations/">http://ethnographymatters.net/blog/2013/04/02/insights-from-network-data-analysis-that-yield-field-observations/</a>, avril 2013.

INNES JON, Usability Testing Is Qualitative Only If You Can't Count,

http://www.uxmatters.com/mt/archives/2011/02/usability-testing-is-qualitative-only-if-you-cant-count.php, février 2011.

JAMES KYLE, *Are you ready for Web Analytics 2.0*?, <a href="http://doteduguru.com/id245-introducing-web-analytics-2-0.html">http://doteduguru.com/id245-introducing-web-analytics-2-0.html</a>, juin 2008.

KAUSHIK AVINASH, *Multiplicity: Succeed Awesomely At Web Analytics 2.0!*, <a href="http://www.kaushik.net/avinash/multiplicity-succeed-awesomely-at-web-analytics-20/">http://www.kaushik.net/avinash/multiplicity-succeed-awesomely-at-web-analytics-20/</a>, 6 novembre 2007.

KOWITZ BRADEN, Should designers trust their instincts — or the data?,

http://www.gv.com/lib/design-instinct-vs-data, 15 janvier 2014.

LAI MICHAEL, When Does Quantity Become Quality? How to navigate big data, <a href="http://uxmag.com/articles/when-does-quantity-become-quality-how-to-navigate-big-data">http://uxmag.com/articles/when-does-quantity-become-quality-how-to-navigate-big-data</a>, 23 septembre 2013.

MA SHANSHAN, Pairing Up Usability Testing with A/B Testing,

http://www.uxmatters.com/mt/archives/2011/07/pairing-up-usability-testing-with-abtesting.php, juillet 2011.

MCLEOD SAUL, Qualitative Quantitative,

http://www.simplypsychology.org/Quantitative%20vs%20Qualitative.pdf, 2008.

MESIBOV MARLI, The right test at the right time,

 $\frac{http://blog.abovethefolddesign.com/2011/09/08/the-right-test-at-the-right-time/,\ 8\ septembre\ 2011.$ 

NOVA NICOLAS, *Ethnomining and the combination of qualitative & quantitative data*, <a href="http://ethnographymatters.net/blog/2013/04/02/april-2013-ethnomining-and-the-combination-of-qualitative-quantitative-data/">http://ethnographymatters.net/blog/2013/04/02/april-2013-ethnomining-and-the-combination-of-qualitative-quantitative-data/</a>, avril 2013.

ROUSI REBEKAH, An uplifting experience – adopting ethnography to study elevator user experience, <a href="http://ethnographymatters.net/blog/2013/04/02/an-uplifting-experience-adopting-ethnography-to-study-elevator-user-experience/">http://ethnographymatters.net/blog/2013/04/02/an-uplifting-experience-adopting-ethnography-to-study-elevator-user-experience/</a>, avril 2013.

SAURO JEFF, 5 Examples of Quantifying Qualitative Data,

http://www.measuringu.com/blog/quantify-qualitative.php, 24 janvier 2012.

SHAH HITEN, Qualitative Questions That Lead to Great A/B Tests,

http://blog.optimizely.com/2014/06/04/qualitative-questions-that-lead-to-great-ab-tests/, juin 2014.

USERLYTICS, *Optimizing A/B User Testing*, <a href="http://www.userlytics.com/blog/optimizing-ab-testing">http://www.userlytics.com/blog/optimizing-ab-testing</a>, 30 septembre 2011.

# MÉTHODOLOGIES D'ENQUÊTE ET D'ANALYSE

ASJES KATHLEEN, The Hidden Benefits of Remote Research,

http://www.uxbooth.com/articles/hidden-benefits-remote-research/, juin 2014.

BERNHEIM BRUSH A.J., *Ubiquitous Computing Fundamentals*, s.l., Chapman and Hall/CRC, 2009, 328 p.

FAURE CHRISTIAN, *Amélioration continue*... *et discontinue*, <a href="http://www.christian-faure.net/2014/02/23/amelioration-continue-et-discontinue">http://www.christian-faure.net/2014/02/23/amelioration-continue-et-discontinue</a>, février 2014.

GOWARD CHRIS, *You should test that !*, John Wiley & Sons, Indianapolis, 2013, 331p. KAUSHIK AVINASH, *Best Web Analytics Tools: Quantitative, Qualitative, Life Saving !*, <a href="http://www.kaushik.net/avinash/best-web-analytics-tools-quantitative-qualitative/">http://www.kaushik.net/avinash/best-web-analytics-tools-quantitative-qualitative/</a>, 19 octobre 2010.

KAUSHIK AVINASH, Context Is King Baby! Go Get Your Own.,

http://www.kaushik.net/avinash/context-is-king-baby-go-get-your-own/, 18 mars 2008. LILLEVALJA JAAN-MATTI, *How to Build a Strong A/B Testing Plan That Gets Results*, http://conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results/, mai 2013.

NAUGHTON JOHN, « The web giants pumping us for data », in : *The Guardian*, <a href="http://www.theguardian.com/technology/2013/sep/01/big-data-corporations-information">http://www.theguardian.com/technology/2013/sep/01/big-data-corporations-information</a>, 1 septembre 2014.

PATEL NEIL, What Spending \$252,000 On Conversion Rate Optimization Taught Me, <a href="http://www.quicksprout.com/2012/10/08/what-spending-252000-on-conversion-rate-optimization-taught-me">http://www.quicksprout.com/2012/10/08/what-spending-252000-on-conversion-rate-optimization-taught-me</a>/, 8 octobre 2012.

PORTER JOSHUA, *The Local Maximum*, <a href="http://52weeksofux.com/post/694598769/the-local-maximum">http://52weeksofux.com/post/694598769/the-local-maximum</a>, juin 2010.

ROHRER CHRISTIAN, When to Use Which User-Experience Research Methods, <a href="http://www.nngroup.com/articles/which-ux-research-methods/">http://www.nngroup.com/articles/which-ux-research-methods/</a>, 12 octobre 2014. RUSONIS SHANA, 5 Essential Steps for Successful A/B Tests Based on Data,

http://blog.optimizely.com/2014/08/07/5-essential-steps-for-successful-ab-tests-based-on-data/, août 2014.

Wolfram Laurissa, *Prepping Open Response Survey Data for Analysis*, <a href="http://us5.campaign-archive2.com/?id=b837b7819f&u=7e093c5cf4">http://us5.campaign-archive2.com/?id=b837b7819f&u=7e093c5cf4</a>, novembre 2012. Young Scott W. H., *Improving Library User Experience with A/B Testing: Principles and Process*,

http://quod.lib.umich.edu/w/weave/12535642.0001.101?view=text;rgn=main, Montana State University, Volume 1, Issue 1, 2014.