



HAL
open science

Croire le vrai, éviter l'erreur

Paul Égré

► **To cite this version:**

Paul Égré. Croire le vrai, éviter l'erreur. Jean-Marie Chevalier; Benoit Gaultier. Le Sens du Réel: Essais en l'honneur de Claudine Tiercelin, A paraître. hal-04187604

HAL Id: hal-04187604

<https://hal.science/hal-04187604>

Submitted on 25 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Croire le vrai, éviter l'erreur

Paul Égré*

Résumé : L'objet de cet essai est de revisiter le débat entre William James et William Clifford touchant la certitude comme norme de la croyance. Alors que Clifford soutient qu'une croyance fondée sur une évidence insuffisante est fautive, James soutient que pour parvenir à croire le vrai quand l'évidence est incertaine, il faut admettre une part d'erreur et de croyances fausses. À l'appui de la thèse de James contre l'agnosticisme supposé de Clifford, je propose un argument inspiré de la théorie de la détection du signal, qui constitue une théorie bayésienne de l'utilité espérée de nos réponses en situation d'incertitude. Cependant, la notion même d'utilité espérée permet aussi d'interpréter différemment l'intuition centrale de Clifford. Ce que Clifford condamne avant tout, ce n'est pas de former une croyance faillible sur la base d'une évidence incertaine, mais de mal évaluer le lien entre cette incertitude et ses conséquences pratiques.

1. Clifford et James sur le doute et la certitude

Dans un texte célèbre sur l'éthique de la croyance (1877), le mathématicien et philosophe William Clifford écrit : « *it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence* ». Il est fautif toujours, partout et pour quiconque, de croire que ce soit ce sur la base d'une évidence insuffisante¹. Le texte de Clifford s'ouvre, de façon dramatique, sur le récit de la mauvaise décision prise par un armateur, celle de mettre à flot pour des migrants un navire dont la solidité lui cause des doutes, mais dont il se dit qu'il a pourtant des chances de réussir la traversée.

Comme l'indique Clifford, la sincérité de l'armateur dans son espérance que la traversée n'échoue pas ne lui ôte nullement la responsabilité de la mort des passagers. Clifford poursuit et affirme en outre que, lors même que le bateau aurait réussi sa traversée sans sombrer ni causer de victimes, l'armateur n'en serait pas moins condamnable, car il aurait seulement eu de la chance. Ce que l'armateur aurait dû faire, c'est prendre toutes les précautions pour que son bateau effectue une traversée sûre.

La maxime de Clifford semble à première vue imparable. Pourtant elle soulève bien des questions : si j'entreprends un voyage, puis-je être certain qu'il n'arrivera pas le moindre imprévu ? Puis-je être certain d'avoir éliminé toutes les causes de doute ? En outre, que faut-il appeler une évidence « suffisante » ? Est-ce quand je suis certain à 100% de la vérité d'une proposition ? Mais existe-t-il de telles propositions, si l'on excepte les vérités purement déductives des mathématiques pour lesquelles la norme de la croyance est une preuve irréfutable ?

* Institut Jean-Nicod (CNRS, EHESS, ENS-PSL). Ce texte en hommage à Claudine Tiercelin est issu d'un exposé donné au Collège de France en novembre 2016 lors du colloque « Certitude et Infaillibilité », organisé dans le cadre de son séminaire, et d'une conférence donnée par la suite à Regensburg fin janvier 2017. Je remercie très vivement Benoit Gaultier et Jean-Marie Chevalier de leurs encouragements et de leurs commentaires sur ce texte, ainsi que les participants du Groupe de recherche en épistémologie (GRÉ). Je remercie Claudine Tiercelin de m'avoir incité à relier mes réflexions au débat entre James et Clifford, et Hans Rott de son invitation à les présenter au sein du département de philosophie de Regensburg. Je remercie également Joëlle Proust et Jakob Süskind pour leurs commentaires sur ce texte. Cette recherche a bénéficié du soutien des programmes ANR-21-ASIA-0003-04 (Hybrinfox), ANR-19-CE28-0019-01 (Ambisense), ANR-17-EURE0017 (FrontCog), et d'une mobilité à l'Université Monash (Programme PLEXUS, Horizon 2020, financement Marie Skłodowska-Curie n°1010866295).

¹ Tout au long du texte, j'utilise le terme *évidence* dans le sens qu'il a en anglais, et qui désigne l'information disponible à un agent connaissant, ou encore « les données probantes » (Gaultier 2018). C'est le sens dans lequel il est utilisé couramment en épistémologie et en psychologie. Voir notamment Engel (2001), qui parle des « données dont on dispose », et Tiercelin (2005), qui utilise également le terme au pluriel (« les évidences disponibles »).

Dans une réponse non moins illustre, « The Will to Believe » (1896), le philosophe et psychologue William James fait apparaître une tension entre deux maximes : « croire le vrai », et « éviter l'erreur ». La principale difficulté de la position de Clifford, nous dit James, est qu'il ne suffit pas de s'abstenir de croire le faux pour croire le vrai. Bien souvent, selon James, la quête du vrai implique de commettre des erreurs. James résume sa position de la manière suivante ² :

« Croyez la vérité ! » ; « Évitez l'erreur ! » sont deux lois dont le contenu est différent ; et le choix que nous faisons entre elles peut conduire à colorer d'un ton tout différent l'ensemble de notre vie intellectuelle. Nous pouvons considérer la poursuite de la vérité comme primordiale et le fait d'éviter l'erreur comme secondaire, ou tout au contraire considérer le fait d'éviter l'erreur comme impératif et laisser à la vérité le soin de prendre sa chance. C'est cette dernière option que Clifford nous exhorte à suivre (...). Ne croyez rien, dit-il, suspendez toujours votre jugement plutôt que de courir le terrible risque de croire le faux pour avoir pris un parti sur la base d'une évidence insuffisante. Mais vous, de votre côté, peut-être estimez-vous que le risque d'erreur est de peu d'importance comparé aux bienfaits de la connaissance réelle, et peut-être êtes-vous prêts à être souvent dans l'erreur dans vos investigations plutôt que d'ajourner indéfiniment toute chance de deviner le vrai. En ce qui me concerne, il m'est impossible de suivre Clifford.

L'affirmation de James suggère par là-même que pour aboutir au vrai, il faut donc accepter, dans certain cas, de croire certaines choses sur la base d'une évidence incertaine. Dans la première partie de cet essai je propose d'examiner la thèse de James à la lumière de la théorie de la détection du signal, qui fournit un cadre très général pour la représentation des décisions en situation d'incertitude (voir Green & Swets 1966, McNicol 2005).

La théorie de la détection du signal traite les erreurs comme une donnée fondamentale en psychologie. Il s'agit d'une approche bayésienne du comportement, fondée sur l'idée qu'un agent rationnel a intérêt à maximiser ses réponses correctes et à minimiser ses réponses incorrectes compte tenu de l'information qui lui est disponible et des utilités positives et négatives attachées aux bonnes et aux mauvaises réponses ³.

L'argument présenté est très simple : dans certaines situations, la maxime de Clifford, entendue comme l'injonction à ne former des croyances que sur la base d'une évidence qui garantit la vérité de la proposition concernée (ce que nous appellerons une *evidence certaine*), empêcherait un agent rationnel de maximiser le nombre de ses croyances vraies.

Toutefois, nous verrons qu'il existe une autre manière d'interpréter la maxime de Clifford, qui s'accorde aussi avec l'épistémologie bayésienne. Ce que Clifford condamne avant tout, ce n'est pas de former une croyance faillible sur la base d'une évidence incertaine (ne garantissant pas la vérité de la proposition concernée), mais de mal évaluer le lien entre cette incertitude et ses conséquences pratiques.

² Pour les extraits des textes de James et de Clifford, je donne la traduction française de B. Gaultier (Ed.) (2018), occasionnellement modifiée en quelques occurrences. Je reproduis ici la traduction de « believe truth! » par « croyez la vérité ! », qui se justifie par les autres occurrences de « truth » dans le texte de James. S'agissant du problème examiné dans cet essai, je préfère toutefois utiliser l'expression « croire le vrai » plutôt que « croire la vérité », notamment parce que cette dernière suggère qu'on doit croire une unique et entière vérité (une interprétation corroborée par plusieurs locuteurs consultés sur cette question), alors que « croire le vrai » enjoint de façon générique à croire différentes propositions vraies, là où elles se présentent. James, incidemment, n'écrit pas « believe the truth! ». La nuance, me semble-t-il, est comparable à celle qu'on peut entendre entre l'injonction « drink water ! » et l'injonction « drink the water ! ».

³ Elle constitue en cela une version opérationnalisée d'une théorie générale de l'utilité épistémique (cf. Pettigrew 2013 pour une introduction). Richard Pettigrew propose lui aussi une justification des maximes de James en termes bayésiens (cf. Pettigrew 2016, 2022). Nos approches sont convergentes, bien que développées de façon indépendante. À ma connaissance, Pettigrew ne mentionne pas la théorie de la détection du signal en tant que telle, mais il s'appuie sur des concepts et résultats généraux issus du même cadre de la théorie de la décision.

2. Admettre l'erreur pour atteindre le vrai

La théorie de la détection du signal semble d'autant mieux fondée à justifier l'intuition de William James qu'il s'agit d'un outil très utilisé en psychologie, et plus particulièrement en psychophysique, où il s'agit de mesurer et de représenter les performances d'agents (humains ou non-humains) dans des tâches de discrimination imparfaite. Ces dernières années, les concepts de la théorie de la détection du signal ont par ailleurs acquis une place centrale dans le domaine de l'apprentissage-machine, rapprochant par là-même l'épistémologie, la psychologie cognitive, et l'intelligence artificielle.

Un exemple d'une tâche de discrimination imparfaite apparaît sous forme de problème dans l'excellent manuel de Don McNicol sur la théorie de la détection du signal⁴. McNicol (2005, p. 15) propose le scénario suivant : on dispose devant vous un paquet de 450 cartes, qui comportent chacune de 1 à 5 ronds noirs marqués de façon aléatoire sur la carte. La distribution des cartes vous est connue : 50 cartes figurent 1 rond, 100 cartes figurent 2 ronds, 150 cartes figurent 3 ronds, 100 cartes figurent 4 ronds, et 50 cartes figurent 5 ronds. Avant de faire passer l'expérience, l'expérimentateur vous informe par ailleurs qu'il va ajouter un rond sur la moitié des cartes de chaque groupe, sans que vous ne puissiez voir lesquelles. L'expérience commence alors. Votre tâche consiste à détecter les cartes qui ont été modifiées, de façon à classer les cartes en deux groupes. Pour chaque carte, vous devez dire si la carte est une carte d'origine (à laquelle aucun rond n'a été ajouté), qu'on appellera une carte N, ou bien une carte modifiée par l'expérimentateur (à laquelle un rond a été ajouté), qu'on appellera une carte S. En pratique, l'agent peut se contenter de dire « S » quand il pense que la carte est une carte S, et de rester silencieux quand il pense que la carte est une carte N.

Quelle est la meilleure stratégie pour classer le maximum de cartes de façon correcte ? La réponse à cette question ne va pas de soi. Imaginons que vous voyiez une carte ne comportant que 1 rond. Vous savez que c'est nécessairement une carte qui n'a pas été modifiée par l'expérimentateur, donc vous pouvez dire avec certitude que c'est une carte N. Il en va de même pour une carte comprenant 6 ronds : vous savez que, nécessairement, cette carte a été modifiée par l'expérimentateur, et donc que c'est donc une carte S. Mais qu'en est-il d'une carte comprenant 2, 3, 4 ou 5 ronds ? Dans les cas intermédiaires, la carte peut aussi bien être une carte d'origine qu'une carte modifiée par l'expérimentateur. Que dire de ces cas-là ?

La théorie de la détection du signal offre une réponse de principe à cette question. Pour la comprendre, il est utile de représenter la distribution des cartes à l'aide d'une figure. L'histogramme de la Figure 1 montre que seules 25 cartes ne comportent plus qu'un seul rond après la modification de l'expérimentateur, puisque sur les 50 de départ la moitié en a reçu un second. De la même façon, parmi les cartes comportant 2 ronds, 50 sont restées inchangées, mais 25 proviennent des cartes à un rond qui ont été modifiées. Et ainsi de suite. Sur l'histogramme, la distribution en trait plein représente donc le nombre de cartes S, et la distribution en pointillés le nombre de cartes N, pour chaque type de carte. Les deux distributions sont les mêmes, mais la distribution S est décalée vers la droite.

⁴ J'ai déjà eu l'occasion d'utiliser cet exemple dans un autre contexte, en lien avec la théorie de l'incertitude liée au vague. Cf. Égré & Bonnay (2010).

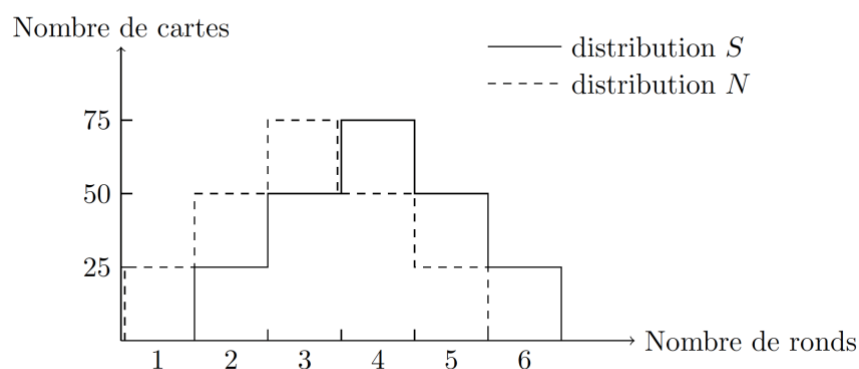


Figure 1: La tâche de classification des cartes (d'après McNicol 2005)

Les cartes comportant 2 à 5 ronds sont celles pour lesquelles vous êtes placé dans une situation d'incertitude, en raison de la compétition entre signal (S) et bruit (N). Il existe différentes stratégies pour classer les cartes, certaines « pures » consistant à classer chaque type de carte de façon uniforme comme N ou S, et d'autres stratégies « mixtes » consistant à donner une réponse probabiliste en fonction du nombre de ronds sur chaque carte.

La stratégie « pure » optimale pour classer le maximum de cartes de façon correcte consiste à déclarer qu'une carte comportant 4 ronds ou plus est une carte S et qu'une carte comportant 3 ronds ou moins est une carte N. En effet, si une carte comporte 3 ronds ou moins, alors on est sûr, en procédant de cette façon, de classer correctement 150 cartes. De même pour les cartes comportant 4 ronds ou plus. Dans le premier cas, on commettra l'erreur de classer 75 cartes S comme des cartes N, et dans le second l'erreur symétrique de classer 75 cartes N comme des cartes S – ce qui signifie que 300 cartes sur 450 auront été correctement identifiées, et 150 mal identifiées. On peut vérifier que pour toute autre frontière de décision (encore appelé « critère »), le nombre de cartes correctement identifiées sera inférieur. Par exemple, si l'on décide que seules les cartes comportant 5 ou 6 ronds sont S, certes on classe mal seulement 25 cartes N, mais on manque aussi 150 cartes S, l'erreur est donc de 175 cartes. L'erreur s'accroît encore dans les autres cas : par exemple, déclarer toutes les cartes ayant 2 ronds ou plus comme S permet de n'en manquer aucune, mais alors on classe mal 200 cartes N.

Qu'en est-il des stratégies mixtes ? Une stratégie mixte spéculative consisterait, en fonction du nombre de ronds sur chaque carte, à choisir de répondre S ou N selon une probabilité reflétant la proportion des cartes S ou N pour chaque type de carte. Par exemple, quand une carte comporte 2 ronds, il est tentant de répondre N deux fois sur trois, et S une fois sur trois, pour refléter le fait qu'exactement deux tiers des cartes en question sont des cartes N. On peut montrer, cependant, qu'une stratégie de ce type (dite de « probability matching », dans laquelle la probabilité reflète la proportion), donnerait un résultat moins bon que la stratégie déterministe mentionnée précédemment, dans laquelle le critère est fixe⁵. Certes, en adoptant une telle stratégie probabiliste, on se donne une chance de classer correctement toutes les cartes, mais en termes d'utilité espérée, le nombre de cartes qu'on peut espérer classer correctement est moindre⁶. La stratégie la meilleure dans la tâche considérée revient donc à tracer une frontière de décision fixe entre les réponses « S » et les réponses « N » en fonction du nombre de ronds sur chaque carte, qui implique de commettre des erreurs.

⁵ Sur la notion de *probability matching*, voir par exemple Koehler & James 2009.

⁶ On peut s'en rendre compte en considérant un cas plus réduit. Supposons qu'il n'y ait en tout et pour tout que six cartes ayant 3 ronds, 4 cartes N et 2 cartes S. Il y a alors 15 distributions de ces 6 cartes. En décidant de répondre « S » ou « N » de façon aléatoire mais en donnant 1 chance sur 3 à « S » et 2 chances sur 3 à « N », alors il y a 1 chance sur 15 de donner la bonne réponse et de classer les 6 cartes, 8 chances sur 15 d'en classer correctement 4, et 6 sur 15 d'en classer correctement 2 : l'utilité espérée est de 48/15 cartes bien classées, ce qui est moindre que le nombre de 4 si on décide de dire « N » invariablement pour chacune des 6 cartes.

Quel est l'intérêt de cette discussion en regard de la maxime de Clifford ? La réponse est simple. Une manière de respecter la maxime de Clifford, interprétée comme le fait James, serait de ne donner une réponse N ou S que pour les cartes pour lesquelles on est certain d'avoir la réponse correcte. Un agent qui suivrait cette maxime devrait donc dire « N » pour les cartes comportant 1 rond, « S » pour les cartes comportant 6 ronds, et s'abstenir de répondre dans les cas intermédiaires. Mais un tel agent « cliffordien » ne pourrait espérer classer correctement que 50 sur les 450 cartes, soit un neuvième, au contraire de l'agent « jamesien » que nous avons décrit précédemment, qui parviendrait à en classer correctement 300 sur les 450, soit les deux tiers.

L'exemple montre qu'il existe certaines situations d'incertitude dans lesquelles il est rationnel de commettre des erreurs pour maximiser ses réponses correctes. Le cas semble d'ailleurs montrer quelque chose de plus fort, à savoir qu'il peut être rationnel de déclarer « cette carte est S » alors même qu'on ne *sait* pas de la carte en question qu'elle est S. Ce constat rend non seulement douteuse la maxime de Clifford, mais il semble aussi contredire la maxime de connaissance de Williamson concernant l'assertion, d'après laquelle on ne doit asserter une proposition qu'à condition de savoir qu'elle est vraie (voir Williamson 2000). La situation, en revanche, s'accorde avec la première partie de la maxime de Qualité de Grice, laquelle enjoint de ne pas dire ce qu'on croit être faux (Grice 1975). En effet, même si l'agent sait qu'il commettra des erreurs, il ne croit d'aucune carte en particulier que sa réponse est fautive (*croissance de re*), tout ce qu'il sait est que certaines de ses réponses seront fausses (*croissance de dicto*)⁷.

3. Le coût des erreurs et l'utilité espérée

Dans la théorie de la détection du signal, la sélection du « critère de décision » par un agent rationnel est fonction non seulement des probabilités de signal et de bruit, mais aussi des gains et des pertes associés aux quatre catégories de réponse qu'on représente usuellement par une matrice dite de confusion, à savoir : l'utilité d'une réponse correcte (*vrai positif*, dire « S » pour « S »), le coût de dire « S » quand en réalité la carte est N (*faux positif*), le coût de dire « N » quand la carte est S (*faux négatif*), et l'utilité de dire « N » quand la carte est N (rejet justifié ou *vrai négatif*).

Par exemple, si l'utilité de dire « S » quand c'est S est de 2 et le coût de dire « N » de -2, tandis que l'utilité de dire « N » quand c'est N est de 1, et le coût de dire « S » de -1, alors on aura intérêt rationnellement à dire « S » dès qu'une carte comporte 2 ronds, sans attendre qu'elle en comporte au moins 4. Autrement dit, on aura intérêt à adopter un critère moins strict que précédemment, qui maximise l'utilité espérée des réponses.

De telles considérations d'utilité permettent toutefois de réévaluer la position de Clifford. Dans son exemple, il suffit de constater que le coût à ne pas prévenir un naufrage, rapporté aux nombres de vies en jeu, est beaucoup plus considérable que le bénéfice à éviter un naufrage en ayant de la chance.

Supposons que l'armateur puisse embarquer 100 passagers à bord et qu'il touche 2 euros par passager embarqué. Supposons qu'il estime à 1/10 les chances que son navire souffre d'un défaut qui entraînera un naufrage. Faisons l'hypothèse que chaque victime d'un naufrage lui coûtera au minimum 10 euros (une valeur naturellement sous-estimée, mais suffisante pour les besoins de notre argument). L'armateur peut choisir de dépenser 50 euros avant le départ, pour vérifier l'état de son navire. Si un défaut est trouvé, il devra remettre à plus tard le fait d'embarquer les passagers et les rembourser. Dans ce cas, on peut représenter les gains et pertes associées à chaque situation comme suit :

Table 1 : Une matrice de décision pour l'armateur de Clifford

	Défaut (p=1/10)	Pas de défaut (1-p=9/10)
Vérification	-50	150
Pas de vérification	-800	200

⁷ On entrevoit ici une manière d'expliquer un autre paradoxe de la croyance, le paradoxe de la préface de David Makinson. Dans l'exemple de Makinson (1965), il est rationnel d'admettre que le livre qu'on a écrit contient au moins une erreur, même si l'on ne croit d'aucune phrase du livre en particulier qu'elle est erronée.

L'utilité espérée qu'il y a à vérifier l'état du bateau est de 130 euros ($= -50 \times 0.1 + 150 \times 0.9$), alors que celle de ne pas le vérifier est de 100 euros ($= -800 \times 0.1 + 200 \times 0.9$). Un armateur rationnel, sous ces hypothèses, a donc intérêt à vérifier l'état de son embarcation⁸. Par voie de conséquence, il est fondé à croire que le bateau mérite une vérification, plutôt que de croire que le bateau ne mérite pas de vérification.

Vu de la sorte, ce qu'indique l'exemple de Clifford n'est pas qu'il est incorrect de former une croyance quand on est incertain, mais plutôt qu'il convient d'agir en prenant en compte l'utilité espérée en situation d'incertitude. De fait, le principal reproche que fait Clifford à l'armateur n'est pas d'agir en étant incertain, mais plutôt de manipuler ses croyances⁹. En effet, l'armateur modifie la perception des probabilités associées à la proposition « le navire est en état de partir en mer ». Clifford décrit avec minutie la manière dont l'armateur met ses doutes de côté :

Certains lui firent part de leurs doutes à propos de la capacité du bateau à partir en mer. Ces doutes le préoccupaient et le rendaient malheureux ; il se disait qu'il devrait peut-être le faire entièrement réviser et rééquiper, même si cela devait lui coûter beaucoup d'argent. Cependant, avant que le navire ne prenne le large, il réussit à laisser derrière lui ces réflexions qui le tracassaient beaucoup. Il s'en remit à la Providence (...) Il rejeta hors de son esprit tous les soupçons mesquins que l'on peut entendre sur l'honnêteté des entrepreneurs et des constructeurs navals. De cette manière, il acquit la conviction sincère et agréable que son navire était parfaitement sûr et en état de partir en mer.

Le tort de l'armateur est donc d'agir comme si la probabilité que son navire n'avait pas de défaut était maximale, *dans une situation où il sait pourtant qu'il y a un doute raisonnable quant à sa vérité*. En manipulant ses croyances, l'armateur se persuade que l'action rationnelle est de ne pas vérifier l'état de son bateau, soit l'action dont l'utilité est en effet la plus haute, mais seulement conditionnellement à l'hypothèse que le scénario le plus probable est celui qui se réalisera (colonne de droite de la Table 1). Comme le résume Clifford :

Il a en effet acquis sa croyance non pas au terme d'une enquête patiente, mais en étouffant ses doutes.

En somme, l'analyse même de Clifford est parfaitement compatible avec une conception de la croyance rationnelle fondée sur la notion d'utilité espérée. Étant donné les utilités et les probabilités de chaque éventualité dans notre exemple, l'armateur *devrait rationnellement* croire que le navire mérite une vérification, et donc ne pas croire qu'il est en état de partir en mer sans vérification. Au contraire, l'armateur fonde sa décision sur le scénario qui lui rapporte le gain le plus élevé, non pas en termes d'espérance d'utilité (au sens mathématique), mais en termes d'utilité conditionnelle. Il s'agit en cela d'un sophisme, causé par l'appât du gain.

⁸ Notons qu'ici les utilités sont considérées du point de vue du seul armateur, mais sachant que la perte de cent vies humaines constitue un dommage beaucoup plus grand en termes d'utilité sociale.

⁹ Sur ce point, voir l'éclairante discussion du texte de Clifford que donne Gaultier (2018), qui montre que Clifford ne prône ni le scepticisme ni l'agnosticisme. Gaultier présente toutefois Clifford comme *anti-pragmatiste*, entendant par-là que nos croyances devraient être imperméables aux considérations d'utilité. L'analyse que je propose diffère de ce point de vue, puisqu'elle replace l'exemple de Clifford dans un cadre utilitariste et pragmatiste. Elle s'accorde en revanche sur le fait que Clifford dénonce un *biais* des considérations d'utilité sur nos croyances rationnelles. Pettigrew (2022, p. 184) envisage quant à lui deux manières de comprendre le raisonnement de l'armateur dans un cadre bayésien. Selon la première, l'armateur choisit de ne pas conditionnaliser sur l'évidence totale dont il dispose, mais sur l'évidence partielle qui favorise la proposition « le navire est en état de partir en mer ». Mon analyse est plutôt que l'armateur modifie ses *priors* de façon à ne pas rechercher plus d'évidence, en regardant l'utilité de ne pas vérifier dans le cas favorable, et en négligeant sa désutilité dans le cas défavorable.

4. Que signifie croire sur la base d'une évidence suffisante ?

Un point tout à fait central dans ce qui précède est que dans chaque scénario, nous devons distinguer la probabilité subjective ou objective de certains événements et la conduite rationnelle à tenir à l'égard de ces événements.

Dans l'exemple de Clifford, ce n'est pas parce qu'une hypothèse est la plus probable *a priori* (« le navire n'a pas de défaut ») qu'on est justifié à croire une proposition qui semble matériellement impliquée par elle (« le navire ne mérite pas de vérification »). De même, dans l'exemple des cartes, il est sensé de croire l'hypothèse que l'évidence rend comparativement la plus probable, mais seulement si le coût d'un faux négatif et d'un faux positif s'équilibrent. Si un faux négatif est beaucoup plus coûteux, alors il est rationnel de commettre plus de faux positifs, et donc de donner plus de poids à l'hypothèse qui serait comparativement moins probable.

Pour qu'une évidence soit *suffisante* à la croyance en une proposition, quand l'évidence en faveur de cette proposition est incertaine, il n'est donc ni nécessaire ni suffisant d'avoir un degré de probabilité élevé dans l'hypothèse, si par là on entend le fait d'avoir un degré de probabilité proche de 1. Au lieu de cela, il faut considérer l'utilité espérée de l'hypothèse comparée à l'utilité des hypothèses concurrentes.

Dans le cas de Clifford, ajoutons que si l'armateur décidait de vérifier l'état de son bateau, pour découvrir qu'il n'a pas en réalité de défaut, il aurait donc aussi commis une erreur, sachant que rétrospectivement son navire pouvait partir en mer sans vérification. Mais cette erreur de commission ne serait pas condamnable au même titre que l'erreur d'omission décrite par Clifford, car sans les mêmes effets néfastes.

Aussi bien James que Clifford, de ce point de vue, défendent des thèses sur la nature de la croyance qui s'accordent avec une épistémologie pragmatiste au sens large, prenant en compte les conséquences de nos actions.

5. Conclusions

L'opposition entre James et Clifford est moins forte qu'elle ne semblait à première vue, puisqu'une approche bayésienne fondée sur la notion d'utilité espérée permet de réconcilier leurs thèses. Dans le cas de James, la théorie de la détection du signal nous permet d'étayer l'affirmation selon laquelle ne croire que ce qui est absolument certain limite notre capacité à reconnaître le vrai comme tel. Dans le cas de Clifford, une analyse fondée sur la notion d'utilité nous montre que si une évidence certaine est par là même suffisante, une évidence peut être suffisante sans être certaine. La faute décrite par Clifford n'est pas de croire une proposition incertaine, mais de manquer de pondérer l'hypothèse la moins probable par le coût plus élevé d'un faux négatif (manquer de vérifier quand il convenait de le faire) comparé à un faux positif (vérifier quand il n'était pas requis de le faire).

Il reste toutefois plusieurs problèmes que nous ne pouvons examiner en détail ici.

L'un d'eux concerne notre connaissance du poids relatif du signal par rapport au bruit en situation d'incertitude. Dans le problème des cartes, les probabilités sont données de façon objective, alors que dans le problème de Clifford elles sont données de façon subjective. Comment nous assurer que les probabilités *a priori* de nos hypothèses sont bien ajustées ? C'est là un problème plus complexe que celui de la conduite à tenir lorsque les probabilités sont données. On peut, naturellement, transposer la question au statut des utilités, et demander ce qui fixe les utilités et de quelle manière en avoir une juste représentation ¹⁰.

¹⁰ Sur cette question, je renvoie à la discussion de la notion d'utilité donnée dans Proust (2012) et Proust (2013). Proust distingue utilités pratiques ou stratégiques et utilités épistémiques, et argumente que les utilités épistémiques sont autonomes relativement aux utilités pratiques. Mon traitement de l'exemple de Clifford ci-dessus ne fait intervenir que des utilités pratiques, pour une proposition dont le contenu est lui-même d'ordre pratique, mais de façon générale il faut concevoir que les utilités pratiques sont pondérées différemment selon la valeur intrinsèque accordée par un agent au fait de donner une réponse correcte ou incorrecte en tant que telle.

Un second problème touche à la manière dont l'agnosticisme aussi est susceptible d'être rationalisé de façon bayésienne dans ce cadre. Supposons qu'un agent ait non pas seulement deux mais trois réponses possibles dans le jeu de cartes, avec désormais une réponse d'incertitude « U ». Il est possible alors de modifier les coûts et bénéfices de chaque réponse de façon à ce qu'un agent rationnel n'affirme S ou N que lorsqu'il est absolument certain, réservant la réponse d'incertitude « U » dans tous les autres cas¹¹. Dans un tel scénario comme dans le précédent, cependant, il semble qu'un agent ne puisse former des croyances ou suspendre ses croyances de façon rationnelle que s'il a une juste perception de la plausibilité que chaque hypothèse confère à ses observations, et de leur poids relatif.

Un troisième problème concerne la question de savoir quelle part revient à la certitude comme norme épistémique. Dans l'exemple du jeu de cartes, nous faisons remarquer qu'un agent peut être justifié à déclarer « cette carte est N » ou « cette carte est S » sans en avoir la connaissance, parce que l'agent cherche à avoir des connaissances aussi correctes que possibles en moyenne. Mais il n'en demeure pas moins une différence épistémologique profonde entre les cas extrêmes des cartes comportant 1 rond et 6 ronds pour lesquelles on peut *savoir* que la carte est S ou que la carte est N de chaque carte prise individuellement, et les cas intermédiaires ne comportant que 2 à 5 ronds, dans lesquels on ne sait d'aucune carte si elle est S ou N à proprement parler. La certitude est une norme beaucoup trop forte pour la croyance rationnelle. Mais elle demeure une norme tout à fait plausible pour la connaissance.

¹¹ Voir Smith et al. 2003 pour une étude de situations de ce type, et Proust (2013) sur l'importance à distinguer, du point de vue métacognitif, les situations de choix forcé entre deux réponses et celles dans lesquelles une réponse d'incertitude est permise. Sur la rationalisation de l'agnosticisme en termes d'utilité épistémique, voir également Pettigrew (2022).

Bibliographie

- Clifford, W. (1877). The Ethics of Belief. *Contemporary review*, 29, pp. 289-309.
- Égré, P. and Bonnay, D. (2010). Vagueness, uncertainty and degrees of clarity. *Synthese*, 174, pp. 47-78.
- Engel, P. (2001). La volonté de croire et les impératifs de la raison: sur l'éthique de la croyance. *Revista da Faculdade de Letras*, 18 (2), pp. 165–176.
- Gaultier, B. (Ed.) (2018). *L'immoralité de la croyance religieuse. L'éthique de la croyance* de William Clifford, suivi de *La volonté de croire* de William James. Nouvelle traduction, introduction et postface de Benoit Gaultier. Agone.
- Gaultier, B. (2018). Comment défendre l'anti-pragmatisme de Clifford à propos des croyances en général et des croyances religieuses en particulier. *ThéoRèmes*. Penser le religieux, (13).
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.
- Grice, P. (1975). Logic and Conversation. In P. Cole and J. L. Morgan (Eds), *Syntax and Semantics: Speech Acts*, volume 3, New York: Academic Press. (William James Lectures, données à Harvard en 1967).
- James, W. (1896). The Will to Believe. In W. James, *The Will to Believe and other Essays*, pp. 1-31, John Wilson and Sons, Cambridge MA.
- Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, 113(1), pp. 123-127.
- Makinson, D. C. (1965). The paradox of the preface. *Analysis*, 25(6), pp. 205-207.
- McNicol, D. (2005). *A primer of signal detection theory*. Lawrence Erlbaum Associates. New edition with a foreword by B.C.J. Moore (First ed. 1972).
- Pettigrew, R. (2013). Epistemic Utility and Norms for Credence. *Philosophy Compass*, 8(10), pp. 897–908.
- Pettigrew, R. (2016). Jamesian epistemology formalised: An explication of 'The Will to Believe'. *Episteme*, 13(3), pp. 253-268.
- Pettigrew, R. (2022). *Epistemic Risk and the Demands of Rationality*. Oxford University Press.
- Proust, J. (2012). The norms of acceptance. *Philosophical issues* 22, pp. 316-333.
- Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. Oxford University Press.
- Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and brain sciences*, 26(3), pp. 317-339.
- Tiercelin, C. (2005). *Le Doute en question*. Éditions de l'Éclat.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.