

UNE RAISON BIEN GARDÉE

Marion Vorms

*Université Paris 1 – IHPST – IUF
marion.vorms@univ-paris1.fr*

MOTS-CLÉS

Biais cognitifs, Irrationalité, Psychologie du raisonnement

F*rench theory*, et psychologie évolutionniste, même combat? Alain Badiou et Steven Pinker, frères d'armes? Certainement pas, mais il est possible, sans être amis, d'avoir un ennemi commun. Le stimulant ouvrage de Pascal Engel prétend ainsi mettre en évidence l'allégeance – assumée ou non – d'un grand nombre d'approches philosophiques contemporaines, de la pensée postmoderne la plus relativiste au naturalisme le plus scientiste, au « Parti de l'anti-raison » (p. 17). Engel nous annonce ainsi dès l'Introduction de son livre qu'« il serait à la fois naïf et injuste de soutenir que [la philosophie française] a le privilège [de l'irrationalisme] » (p. 22). Selon son diagnostic, de très nombreuses théories, tant philosophiques que scientifiques, pour se réclamer d'une certaine rationalité, reposent en fait sur un renoncement à la raison.

Si l'on refuse, avec Pascal Engel, l'idée d'une raison plurielle et contextuelle, force est de constater que la défense du rationalisme, parce qu'il est attaqué de toutes parts, se doit d'être protéiforme. Du rationnel au raisonnable, des raisons internes ou externes à la Raison – pratique ou théorique –, se dégage de ce parcours qui, de l'aveu même de l'auteur, ne suit pas « un ordre des raisons » (p. 26), l'idée d'une exigence, celle de ne tomber dans aucun des multiples travers décrits dans ce livre. Est-il possible, toutefois, d'éviter tous ces écueils? La raison, qui n'est jamais explicitement définie, est-elle autre chose que ce à l'ombre de quoi l'on peut se trouver?

Dans ce qui suit, je me concentrerai sur la forme d'anti-rationalisme que Pascal Engel dénonce chez les « sources savantes » (p. 22) qui étudient empiriquement les raisonnements des humains et apportent « la mauvaise nouvelle » (*ibid.*) de leur irrationalité manifeste et irrémédiable. Que peut-il y avoir de commun entre le rejet de la rationalité comme norme et le constat empirique des dysfonctionnements de la faculté humaine de raisonner? « Optimistes

béats » et « pessimistes ravis » (p. 312) le sont au regard d'une norme, qu'ils endossent donc. Ils incarnent toutefois, selon le diagnostic de Pascal Engel, un rationalisme « hémiplegique » (p. 314) — un rationalisme « obscur » (p. 316) qui, « sous couvert de démystifier la raison, [...] fait de celle-ci le produit de forces causales qu'elle ne peut pas maîtriser » (p. 316) et renonce à l'exigence dictée par le « quatrième pilier de la raison » selon lequel « les raisons doivent être au moins internes » (p. 326). L'examen du diagnostic de Pascal Engel – dont il convient de dire d'emblée que je partage l'essentiel – et la confrontation des exigences de ce quatrième pilier avec le projet même d'étudier expérimentalement le raisonnement, me semblent toutefois révéler une tension qui me conduit à soumettre une question à l'auteur du *Manuel rationaliste de survie* : la psychologie expérimentale du raisonnement peut-elle être rationaliste, ou bien la démarche au fondement de toute enquête empirique sur la rationalité implique-t-elle de renoncer à une certaine idée de la raison, seule digne du rationalisme de Pascal Engel ?

1 LA « MAUVAISE NOUVELLE »

Le huitième et dernier chapitre du *Manuel*, intitulé « La raison claire obscure », est consacré à « quelques formes ambiguës de rationalisme issues de travaux inspirés par les sciences cognitives et la psychologie évolutionniste qui nient l'existence d'une capacité humaine telle que la raison » (Introduction, p. 27). Notons que Pascal Engel englobe dans une seule et même catégorie, qu'il appelle les « théories psychologiques de la rationalité » (p. 329), les travaux issus de la psychologie sociale, de la psychologie cognitive du raisonnement, de la psychologie évolutionniste ou encore de la psychologie du jugement et de la prise de décision. Bien qu'elles convoquent des explications théoriques et reposent sur des paradigmes expérimentaux très différents, toutes ces approches ont en commun, d'après lui, de nous apporter la « mauvaise nouvelle » selon laquelle « les humains ne sont pas des animaux rationnels, [...] font régulièrement des erreurs grossières de logique et sont soumis à toutes sortes de biais qui conduisent à des jugements et à des actions irrationnels » (p. 22).

Plusieurs interprétations de ce constat sont données dans la littérature. Selon les plus pessimistes, « l'ensemble des compétences rationnelles que l'on attribue couramment aux êtres humains » (p. 318) est menacé par l'omniprésence de nos infractions patentées aux règles de la logique, du calcul des probabilités et de la théorie de la décision¹. Plus optimistes, d'autres affirment que l'évolution nous a façonnés pour obéir à une forme de rationalité « écologique »² : pour enfreindre les règles de la rationalité classique, nous n'en sommes pas moins adaptés à notre environnement et nos « biais » ne

1. Pascal Engel semble attribuer cette position à Gérald Bronner (2013).

2. Le plus célèbre tenant de cette position est Gerd Gigerenzer (2000).

sont que des effets secondaires d'« heuristiques » qui conduisent la plupart du temps à des résultats désirables³.

Ces différentes interprétations des déviations du comportement humain vis-à-vis des normes classiques de rationalité sont autant de manières, pour Pascal Engel, de renoncer à la raison. Même quand elles rendent « raison » de nos comportements en montrant qu'ils se conforment d'une certaine manière à un certain type de norme rationnelle, cette supposée conformité se réalise à notre insu. Ce « rationalisme de surplomb » (p. 321) ne nous accorde en aucun cas l'accès aux raisons, qui constitue pourtant l'un des piliers fondamentaux du rationalisme tel que Pascal Engel l'entend : « c'est le psychologue qui a raison, pas nous » (*Ibid.*).

2 UN RATIONALISME INHUMAIN

Qu'il s'agisse de s'en désoler ou de tenter d'y remédier, les travaux des « psychologues du raisonnement et la psychologie sociale [...] ont en un sens confirmé le pessimisme des moralistes sur l'impuissance de la raison » (p. 317). Si les naturalistes contemporains incarnent bien une forme de rationalisme – puisque ce constat empirique d'irrationalité se fait sur fond de l'acceptation initiale (au moins implicite) de normes du raisonnement correct – il s'agit d'un rationalisme proprement inhumain. Contrairement aux moralistes classiques, mais aussi à Hume qui « savait ce qu'est la vertu » (p. 351), le péché de leurs héritiers contemporains est en effet d'avoir abandonné l'idée d'une éducation à la raison. Qu'ils aient ainsi perdu foi dans les pouvoirs de la raison est révélé par le fait qu'ils « nous proposent soit de prendre l'irrationalité humaine comme un fait irrémédiable, soit de ne tout simplement pas user de notre raison pour y remédier » (p. 350).

Ce rationalisme contemporain, s'appuyant sur la science pour déplorer les dysfonctionnements, voire l'inexistence de la faculté de raison – ou encore célébrant de manière naïve son avènement⁴ –, est un rationalisme sans humanisme. Ce scientisme non réflexif se révèle en particulier dans le paternalisme à l'œuvre dans la « douce police des esprits » (p. 351) en laquelle résulte le recours aux « *nudges* » : la psychologie sociale et comportementale ayant supposément identifié des « biais » et autres mécanismes par lesquels

3. Une autre voie est présentée et critiquée par Engel, celle de Dan Sperber et Hugo Mercier (2017) selon lesquels la faculté de raison telle qu'on la conçoit classiquement n'est pas tant dysfonctionnelle qu'inexistante, la raison humaine étant un outil de persuasion, qui a évolué de manière à nous permettre de justifier nos croyances et nos actions auprès de nos congénères, bien davantage qu'elle n'est un instrument destiné à parvenir de manière solitaire à des croyances exactes et des décisions justes. La critique de Sperber et Mercier par Engel mériterait une réponse séparée, que je n'entreprendrai pas ici. Notons toutefois que leur approche repose en partie sur une critique de certains travaux expérimentaux analogue à celle que je propose dans ce qui suit.

4. Engel attribue cette position à Steven Pinker (2018).

nous serions dévoyés des chemins de la rationalité, il s'agit d'utiliser ces ressorts pour inciter les sujets à adopter les comportements les meilleurs pour eux – ceux qu'ils adopteraient *s'ils étaient rationnels* – sans qu'il leur soit nécessaire (ni possible) d'accéder aux véritables raisons (celles qui rendent ces comportements rationnels d'un point de vue externe). C'est là le comble du dévoiement purement instrumental de la raison : seules les conséquences pratiques comptent, peu importent le processus intellectuel et la compréhension par les sujets des raisons véritables de leur comportement rationnel en apparence :

Cette tradition entend donner les mécanismes de la croyance, mais aussi ceux de l'illusion de justifier celle-ci. C'est une forme de rationalisme hémiplégique qu'on peut appeler « obscurantisme subtil », ou « raison claire obscure » parce que, dans nombre de ses versions, elle néglige les fonctions essentielles de la raison, et en définitive conduit à traiter les humains comme irrationnels, sans espoir de guérison. (p. 314-315)

Non seulement le « rationalisme de surplomb » du psychologue n'est pas satisfaisant, mais il n'est pas tenable. Si les humains ne sont pas rationnels, si la raison est sans prise sur leurs inférences, simples enchaînements causaux sans lien avec l'ordre des raisons, qui peut l'être, et que vaut la raison ?

Une réponse classique à laquelle le rationaliste pourrait être tenté de se raccrocher, celle de L. Jonathan Cohen (1981), consiste à souligner que le constat empirique de l'irrationalité humaine ne peut être fait que sur fond de l'acceptation du postulat d'une rationalité minimale. En conséquence, la thèse d'une irrationalité massive est impossible à tester empiriquement au sens où elle n'est tout simplement pas compatible avec le rationalisme de principe nécessaire pour poser la question de la conformité des comportements à la rationalité – et pour leur évaluation au regard de cette norme. Cette condamnation de principe de toute tentative pour tester empiriquement la thèse de l'irrationalité humaine semble insuffisante aux yeux de Pascal Engel, car en faisant de la rationalité humaine « un *a priori* ou une tautologie » (p. 319), elle ne fournit aucune explication de « la persistance de certains types d'erreur et de biais, cognitifs et motivationnels » que, selon lui, « les travaux des psychologues montrent de manière robuste » (*ibid.*).

En effet, si Pascal Engel en critique les diverses interprétations, il ne remet pas en question la validité du constat empirique de la massivité et de l'omniprésence des déviations du comportement des humains vis-à-vis des normes de la rationalité. Je serai beaucoup moins charitable que lui à cet égard. Il existe en effet des raisons *empiriques* de remettre en cause la thèse de l'irrationalité humaine et de contester ce point que Pascal Engel semble considérer comme largement établi, à savoir que « la psychologie expérimentale » a effectivement mis au jour de manière robuste des déviations significatives et systématiques à l'égard de normes rationnelles bien définies.

3 QUELQUES DOUTES SUR LA ROBUSTESSE DES RÉSULTATS EXPÉRIMENTAUX ACCRÉDITANT LA THÈSE DE L'IRRATIONALITÉ

Peut-on véritablement affirmer, comme Pascal Engel le concède volontiers à ses adversaires, qu'une « vaste littérature a montré que, sur des tâches élémentaires de raisonnement logique, les humains commettaient des erreurs récurrentes et tenaces, qu'ils étaient sujets à des illusions cognitives et à des biais, et que les affects et les émotions influaient durablement sur leurs capacités cognitives et leurs décisions » (p. 317) ? Il n'est pas question de nier en bloc l'importance des travaux de psychologie sociale, de psychologie du raisonnement et de l'étude expérimentale du jugement et de la prise de décision qui ont mis au jour des infractions aux règles du raisonnement correct. Mais plusieurs points méritent d'être soulignés, que je ne ferai que mentionner ici car ils ont pour la plupart fait l'objet de critiques détaillées de la part de certains psychologues du raisonnement⁵ et psychologues sociaux⁶, critiques dont j'ai moi-même tenté de restituer certains aspects dans un article récent (Vorms, 2021).

D'abord, les différents travaux constituant la « vaste littérature » visée dans l'affirmation reproduite ci-dessus méritent d'être clairement distingués, tant du point de vue des phénomènes qu'ils prétendent mettre au jour, des normes de rationalité dont ces phénomènes semblent manifester une infraction, que des explications théoriques qu'ils en donnent. En effet, les « biais » au cœur de la tradition dite des « biais et heuristiques » engagée par Tversky et Kahneman (1974) dans le domaine de l'étude du jugement et de la prise de décision en situation d'incertitude ont peu en commun avec les innombrables biais supposément mis au jour par la psychologie sociale et dont l'inventaire est relayé au grand public par les « zététiciens »⁷ et autres vulgarisateurs⁸. Quand les premiers correspondent à des infractions manifestes aux règles du calcul des probabilités, les seconds sont plus difficiles à distinguer de simples erreurs de raisonnement ou d'un mauvais traitement des informations, dont ils désignent tantôt les causes, tantôt les effets observables. Alors que les premiers se voient attribuer des causes d'ordre cognitifs, les seconds sont plus souvent expliqués en termes motivationnels.

En outre, comme l'expliquent Hahn et Harris (2014), pour que la mise au jour de déviations vis-à-vis de modèles de rationalité permette de conclure à l'irrationalité, il faut non seulement que les normes au regard desquelles les comportements sont étudiés soient clairement énoncées (ce qui n'est pas

5. Voir en particulier Hahn et Harris (2014).

6. Voir Funder (2000) ; Krueger et Funder (2004).

7. Voir par exemple la chaîne YouTube *La Tronche en Biais*.

8. Le « codex des biais cognitifs » (<https://inertian.wixsite.com/codexbiais>), destiné à aiguïser l'esprit critique et à nous mettre en garde contre ces pièges que nous tendent nos propres fonctions cognitives en recense près de 200.

toujours le cas en psychologie sociale où elles sont bien souvent implicites, intuitives et inhérentes au montage expérimental lui-même⁹), mais encore que les déviations en question soient significatives et se produisent systématiquement dans des contextes variés. Comme Hahn et Harris le montrent, cela n'est le cas ni en psychologie sociale (où à pratiquement chaque biais correspond un biais opposé et où les effets observés dépendent du contexte expérimental au point qu'ils peuvent être annulés, voire s'inverser, dans certaines situations) ni dans la tradition des « Heuristiques et biais »¹⁰. La robustesse fait défaut aux phénomènes en question pour que l'on puisse en tirer une théorie de l'irrationalité humaine au pouvoir explicatif et prédictif substantiel. Comme le suggèrent Krueger et Funder (1982), plutôt que d'accréditer une conclusion pessimiste sur les faibles compétences rationnelles des humains, les infractions à la rationalité que la psychologie cognitive et sociale a su mettre au jour devraient être interprétées comme des exceptions à une théorie plus générale de l'inférence humaine, dont elles pourraient nous ouvrir les portes, comme les illusions d'optique ont permis de comprendre les règles générales dont elles sont des « défauts » significatifs mais marginaux au regard de la remarquable sophistication de notre système visuel¹¹.

Certes, l'affirmation de Pascal Engel mentionnée en début de section n'est pas le cœur de son propos, lequel semble d'ailleurs plutôt renforcé par cette remise en cause du soutien empirique dont jouirait la thèse de l'irrationalité. Il ne concède la robustesse des résultats que pour mieux souligner son désaccord avec les interprétations qui en sont données :

Certains des résultats expérimentaux des psychologues peuvent venir d'erreurs de performance ou d'effets de perspective induits par le format des tests, mais il n'y a aucune raison de ne pas accepter ces résultats. Devons-nous pourtant accepter les interprétations qui en sont données ? (p. 321)

Toutefois, cette concession me semble problématique au regard même du propos de l'ouvrage. Rappelons qu'il s'agit ici pour Pascal Engel de lutter contre une certaine forme d'irrationalisme, ancrée dans un scientisme aveugle consistant à tirer directement d'un ensemble (considérable) de résultats expérimentaux un verdict général et irrévocable sur la rationalité humaine. Il pose pour ce faire une frontière nette entre les résultats de « la science », qu'il n'entend pas contester, et les interprétations que l'on peut en

9. Sur ce point, voir Vorms (2021, 575) ; Hahn et Harris (2014).

10. C'est ce qui fait conclure à David C. Funder (2000, 674) que « la vaste littérature sur les heuristiques et les biais ne comporte guère davantage qu'un ensemble de casse-têtes que la plupart des gens échouent à résoudre quand quelques-uns – sans instructions spécifiques et en dépit de tout – y parviennent néanmoins » (ma traduction).

11. Notons qu'une telle étude était un des objectifs initiaux des travaux sur les biais de Kahneman et Tversky, qui constatent néanmoins que ce projet a été dévoyé : « bien que les erreurs de jugements ne constituent qu'une méthode permettant d'étudier certains processus cognitifs, la méthode est devenue une partie significative du message » (1982, 124, ma traduction).

donner. Cette frontière entre résultats expérimentaux et interprétations mériterait toutefois d'être précisée. Telle que Pascal Engel semble la concevoir, établir cette distinction revient à faire tout à la fois trop et trop peu confiance à la science. Trop, au sens où la psychologie expérimentale est une science jeune, dont les fondements théoriques et les paradigmes expérimentaux sont si mouvants et divers que l'on ne peut pas attendre des travaux même les plus sérieux et méthodologiquement impeccables de nous délivrer des vérités susceptibles d'être acceptées religieusement ou en tout cas avec la même confiance que celle que l'on porte aux théories physiques les plus solides¹². Trop peu confiance aussi, dans la mesure où il semble écarter la possibilité pour la science empirique du raisonnement d'une forme de réflexivité qui lui est pourtant indispensable – moins encore qu'aucune autre science empirique, l'étude du raisonnement ne peut se contenter d'être purement descriptive.

C'est d'ailleurs un verdict d'incapacité de principe à enquêter empiriquement sur le raisonnement que Pascal Engel semble ainsi prononcer à l'égard de la psychologie expérimentale :

La conception du jugement qui sous-tend la méthodologie en psychologie sociale et en sciences cognitives repose sur l'idée que les jugements sont des réponses spontanées des sujets à des questionnaires sur la base d'« intuitions ». Il n'y a rien à objecter à cette méthodologie, *qui est celle de tout travail sérieux en psychologie*. Elle produit des diagnostics d'erreurs et de biais qui abondent dans la littérature et sur la base desquels les théoriciens de la rationalité produisent leurs propres verdicts. Mais cette méthodologie n'a-t-elle pas elle-même ses biais ? Elle prend les réponses aux tests comme des « jugements » au sens de réponses intuitives spontanées. Toutefois, dans la vie réelle, on ne juge pas simplement en réponse à des tests mais selon une dimension diachronique, en intégrant la réflexion. (p. 348, je souligne)

12. Cette foi pratiquement aveugle en la science « sérieuse » me semble particulièrement bien illustrée par l'attitude des « zététiciens », qui tout en se faisant chantres de l'esprit critique, en manquent parfois vis-à-vis des données scientifiques qu'ils relaient. De même, le rejet du post-modernisme et du relativisme s'accompagne souvent d'une forme de sacralisation de l'analyse statistique, comme si elle était davantage qu'un outil (Ainsi Gérald Bronner, invité des *Chemins de la philosophie* sur France Culture le 3 mars 2022, mettait-il au défi ceux qui voudraient l'accuser de laisser l'idéologie influencer sa pensée scientifique : « qu'on me montre exactement à quel moment j'ai refusé l'état de la science [...], quand est-ce que j'ai refusé d'admettre des données statistiques robustes avec un appareil statistique comme la régression multivariée »). Je ne crois pas me tromper en disant que Pascal Engel porterait le même regard que moi sur ce genre d'attitude. Pourtant, de manière ironique, alors même que son ouvrage vise à s'attaquer à ces formes de scientisme aveugle, il me semble que la frontière qu'il entend dresser entre résultats expérimentaux inattaquables et interprétations implique une conception de la science qui peut prêter le flanc à cette critique.

Est-ce à dire, en miroir de l'analyse de L. Jonathan Cohen selon qui toute entreprise de test empirique de la rationalité repose sur un présupposé de rationalité, que l'entreprise même d'étudier expérimentalement le raisonnement est condamnée (puisque c'est « la méthodologie de tout travail sérieux en psychologie ») à renoncer à l'idée d'une forme de rationalité, celle qui intègre la dimension diachronique de nos raisonnements, laquelle seule est susceptible de nous fournir nos raisons véritables? Pour le dire autrement, faut-il conclure des propos de Pascal Engel que le renoncement à la raison des interprétations contemporaines des résultats de la psychologie expérimentale n'est pas tant celui d'une certaine manière d'expliquer les phénomènes mis en évidence, qu'une propriété intrinsèque à toute démarche expérimentale visant « à tester empiriquement les capacités humaines d'inférence, de jugement et de décision » (p. 317)?

Que Pascal Engel soit ou non prêt à admettre une telle conclusion, il est simplement faux que tous les travaux expérimentaux visant à évaluer la rationalité des raisonnements humains concluent à la négative. Non seulement il existe, comme je l'ai montré, des « raisons de ne pas accepter [les] résultats » sur lesquels s'appuient le verdict d'irrationalité, mais les critiques qui forment ces raisons émanent de psychologues cognitifs qui fournissent tout à la fois une autre « interprétation » et produisent des expériences conduisant à des résultats différents de ceux censés accréditer la thèse de l'irrationalité. Je propose d'en dire quelques mots dans ce qui suit; restera alors à déterminer si l'épistémologie naturaliste rationaliste qui se dégage de ces travaux satisfait aux exigences du rationalisme selon Pascal Engel.

4 LA « MÉTHODOLOGIE DE TOUT TRAVAIL SÉRIEUX EN PSYCHOLOGIE » ?

Il existe une approche en psychologie du raisonnement qui repose sur le double postulat selon lequel, d'une part, la bonne norme du raisonnement, tant inductif que déductif, est la norme bayésienne, et, d'autre part, la cognition humaine s'apparente à la résolution de problèmes inférentiels de type probabilistes¹³. Sans entrer dans le détail de l'appareil théorique ni des protocoles expérimentaux de ce vaste programme de recherche, je me contenterai ici de soulever quelques questions quant à la place qu'il occuperait dans le paysage dressé par Pascal Engel. L'approche bayésienne propose tout à la fois une relecture de certains résultats expérimentaux sous un jour

13. Le programme bayésien en psychologie du raisonnement, baptisé « nouveau paradigme » par David Over (2009) a été formulé à l'origine par Mike Oaksford et Nick Chater (voir Oaksford et Chater (2007), (2020)) pour le raisonnement déductif, et fait actuellement l'objet d'un développement remarquable et d'une application à tous les types de raisonnement et à l'argumentation par Ulrike Hahn et ses collègues (voir par exemple Hahn et Oaksford (2006), Corner et Hahn (2013)). Voir Vorms (2021) pour une présentation plus détaillée.

plus charitable à l'égard de la raison humaine¹⁴ et des expériences nouvelles permettant de mettre au jour la conformité, au moins qualitative, des comportements inférentiels des agents avec la norme bayésienne.

Cette approche est-elle une illustration supplémentaire d'une manière de rationaliser *a posteriori*, au moyen d'une conception différente (par exemple « écologique ») de la raison, des comportements qui semblent irrationnels au regard des normes classiques? Un tel jugement est indéfendable, dans la mesure où l'approche bayésienne tend au contraire à montrer, que, au moins qualitativement et à une certaine finesse de grain, les réponses des sujets sont conformes aux prédictions des modèles bayésiens¹⁵.

Cette voie ne me semble relever ni d'un « optimisme ravi » célébrant l'avènement de la Raison ni d'un « pessimisme béat » se désolant de l'inadaptation humaine. Pessimiste, elle l'est, au sens où il ne s'agit pas pour les psychologues bayésiens de nier l'existence de phénomènes marquant l'échec flagrant de la rationalité dans la vie réelle et ses conséquences délétères pour la paix sociale, la santé publique et la démocratie. Mais plutôt que d'expliquer ces dérives par le postulat d'une irrationalité fondamentale, elle cherche à comprendre ce qui, dans ces comportements, est rationnel. Elle incarne ce faisant une idée qui traverse selon moi le livre de Pascal Engel, à savoir qu'une raison qui ne serait aucunement en mesure de guider les inférences humaines (ou que les humains seraient proprement incapables de mettre en œuvre), serait à peu près sans valeur. Reste que l'approche bayésienne en psychologie du raisonnement est un programme expérimental. Il n'est pas – et ne peut pas être – question dans ce paradigme de se prononcer sur le contenu des représentations mentales et les processus inférentiels effectivement suivis par les sujets : dans les termes de David Marr (1982), l'approche bayésienne se situe au niveau computationnel, celui du calcul, et non au niveau de sa mise en œuvre (*implementation*) concrète. En conséquence, elle ne permet pas de saisir l'accès aux raisons par les sujets. Mais quelle démarche expérimentale le pourrait? Faut-il en conclure que cette approche, comme toute étude expérimentale du raisonnement, opère une réduction de la Raison à une rationalité strictement instrumentale?

14. Ainsi par exemple Harris et Hahn (2011) montrent-ils que des agents bayésiens qui se trouveraient dans la situation de tâches classiques comme celles de Weinstein (1980) supposées révéler un biais d'optimisme, apparaîtraient eux-mêmes déraisonnablement optimistes.

15. En faisant ainsi une place à la diversité des « priors », les modèles bayésiens permettent de prédire des différentiels de réponse qui, dans la tradition de psychologie sociale, étaient interprétés comme les indices du caractère pathologique des sujets. Ainsi en va-t-il par exemple de la polarisation des opinions, que Lord *et al.* (1979) considèrent comme une preuve de l'« assimilation biaisée » des informations, et que le modèle bayésien d'Olsson (2013) permet de prédire. Sur ce point aussi, voir Hahn et Harris (2014) et Vorms (2021). À propos de l'objection de circularité et d'auto-justification que l'on peut faire à la démarche bayésienne et de l'articulation du normatif et du descriptif, voir Vorms (2021, 580-1).

5 DESCARTES OU SPINOZA ?

Face aux constats expérimentaux de l'omniprésence des infractions humaines à la rationalité, Pascal Engel propose, plutôt qu'une conclusion pessimiste fataliste ou qu'une solution paternaliste comme le recours aux « *nudges* », d'en appeler à une éducation de la raison – à une « réforme de l'entendement et du jugement » (p. 349). Nous autres humains sommes des êtres doués de raison et cela se traduit, sinon par de bonnes performances dans les tâches que nous soumettent les psychologues expérimentaux, par notre capacité de réflexion et de compréhension. Par la réflexion, nous avons un accès interne aux raisons, qui nous permet de corriger nos erreurs :

De même dans les phénomènes de persévérance de la croyance : les faits, même quand on nous les montre, ne nous font pas changer d'avis. Mais si nous accédons aux raisons qui nous ont fait juger en premier lieu et les comprenons, nous pourrions changer d'avis. Plus banalement, on pourrait dire que la raison suppose la capacité de compréhension. Sans celle-ci, pas de possibilité de critique de soi-même et des autres. (p. 332)

Ces considérations n'ont de toute évidence pas vocation à être traduites en des thèses psychologiques que l'on pourrait tester empiriquement. Quelles prédictions expérimentales, au demeurant, pourrait-on en tirer ? Tenter de répondre à une telle question reviendrait sans doute aux yeux de Pascal Engel à « mécaniser » les raisons – à en faire des causes. Toutefois, les propos rapportés ci-dessus ne sont pas strictement normatifs, au sens où ils affirment bien quelque chose à propos de la faculté de raison des humains – de leur capacité à la mettre en œuvre, et pas seulement des normes qu'ils devraient suivre s'ils étaient rationnels. Ils cohabitent néanmoins sans heurt avec le constat expérimental de nos égarements irrationnels puisqu'ils *ne parlent pas de la même chose*. Quelle marge de manœuvre, dès lors, reste-t-il au psychologue pour étudier le raisonnement humain sans renoncer à la raison au sens où l'épistémologue l'entend ?

La psychologie expérimentale peut, selon moi, aller plus loin dans le dialogue avec l'épistémologie. Je me propose d'illustrer cette thèse par un exemple, opportunément appelé par les dernières pages de l'ouvrage (p. 349 *sqq*), qui mentionnent l'opposition entre deux conceptions du jugement et de la formation des croyances, attribuées respectivement à Descartes et à Spinoza. Alors que la conception cartésienne admet la possibilité de la suspension du jugement, la conception spinoziste suppose que la seule compréhension d'un énoncé implique son assentiment, lequel est donc préalable à tout examen de la plausibilité de cet énoncé. La conception spinoziste implique une forme d'asymétrie entre l'acceptation (immédiate et involontaire) d'un énoncé et son rejet, qui ne peut se faire qu'au terme d'un processus de réflexion. Pour Engel, la psychologie contemporaine est spinoziste – un spinozisme qui aurait renoncé au projet d'une réforme de l'entendement. C'est indéniable, si l'on en croit l'influence remarquable qu'a eue sur la psychologie sociale le travail

du psychologue Daniel T. Gilbert (1991) qui prétend avoir prouvé expérimentalement le caractère « spinoziste » de notre système cognitif. En très bref, au moyen d'un protocole d'une ingénieuse simplicité, Gilbert *et al.* (1990) montrent que des sujets à qui l'on présente des énoncés qui sont tantôt suivis d'une confirmation, tantôt d'une dénégation, ont tendance, pour peu qu'ils soient distraits par une autre tâche, à garder en mémoire l'énoncé en le tenant pour vrai. Gilbert et ses collègues en concluent au caractère automatique et premier de l'acceptation, jugée indissociable de la compréhension même de l'énoncé. Leurs expériences atteignent des records de citation dans les travaux contemporains portant sur la persévérance des croyances, l'influence continuée des fausses informations après dénégation, la force des rumeurs, et plus généralement tout un ensemble de phénomènes que l'on attribue de près ou de loin à ce fameux et redoutable « biais de confirmation »¹⁶.

On peut toutefois remettre en question les résultats de Gilbert et ses collègues, non pas seulement en leur donnant une autre interprétation, mais en infirmant, expérimentalement, l'hypothèse qu'ils sont censés confirmer. Ainsi, dans une étude récemment publiée (Vorms *et al.*, 2022), nous avons montré qu'en ajoutant une variable de plausibilité au protocole de Gilbert *et al.* (1990), on parvenait à des résultats incompatibles avec la thèse de Gilbert : pour les énoncés peu plausibles, le schéma de Gilbert *et al.* s'inverse¹⁷. Ainsi, dans notre étude, la tendance des participants à accepter ou rejeter des énoncés nouveaux est parfaitement alignée avec leur plausibilité initiale (testée expérimentalement). Reste alors, pour expliquer les résultats de Gilbert (et nos résultats pour les énoncés plausibles) à convoquer une hypothèse de confiance épistémique par défaut – dont l'épistémologie du témoignage¹⁸ a depuis longtemps mis en évidence le caractère non seulement raisonnable mais indispensable à la vie sociale et à l'existence des communautés linguistiques¹⁹.

16. Voir par exemple Kessler, Braash et Kardash (2019) ; Pennycook, Cheyne, Barr, Koehler et Fugelsang (2015) ; Risen (2016). L'influence de la théorie de Gilbert dépasse la psychologie et Eric Mandelbaum (2014) lui a donné une dignité philosophique en en faisant le fondement de sa théorie de la croyance. Kahneman (2003) présente quant à lui les travaux de Gilbert comme un soutien en faveur de sa théorie des deux systèmes.

17. Précisons que les énoncés que nous avons proposés aux sujets leur étaient inconnus, sans quoi l'on pourrait vouloir expliquer nos résultats par l'hypothèse selon laquelle les sujets répondaient en fonction de leurs connaissances préalables. L'hypothèse selon laquelle ils répondaient en fonction de la plausibilité évaluée après coup est elle-même évacuée par certains aspects de nos résultats qui révèlent un authentique apprentissage pendant la première phase de l'expérience (voir Vorms *et al.*, 2022, Discussion).

18. Voir par exemple Coady, 1973.

19. Street et Kingstone (2017) font aussi cette hypothèse, en soulignant notamment l'importance de la fiabilité de la source. Sperber *et al.* (2010) quant à eux insistent sur l'absence de pertinence des énoncés proposés aux sujets par Gilbert *et al.* (1990), lesquels n'ont aucune raison de mettre en œuvre les mécanismes de « vigilance épistémique » dont ils sont pourtant dotés. Enfin, Mercier (2017) insiste sur l'existence d'un mécanisme de « contrôle de la plausibilité »

Une telle étude – et de nombreuses autres semblables – produisent-elles des critiques des résultats ou des interprétations nouvelles ? La distinction me semble difficile à faire, mais c'est bel et bien dans et par l'expérimentation que les preuves de l'irrationalité humaine sont ici battues en brèche. Cela permet d'apporter un autre type de réponse que celui cité au premier paragraphe de cette section : notre supposée cécité aux faits, ce « biais de confirmation » qui nous rend imperméables aux informations contrariantes, ne serait pas tant une résistance irrationnelle devant s'expliquer exclusivement par nos motivations ou nos limitations cognitives, qu'un effet de la dynamique subtile entre plausibilité des contenus (pour nous – au regard de nos croyances préexistantes), crédibilité des sources (*idem*) et enjeux d'un changement de croyance – autant d'aspects de notre situation épistémique qui sont façonnés par notre expérience passée.

Cette approche en psychologie du raisonnement, charitable et selon moi authentiquement rationaliste, me semble compatible avec une conception du raisonnement comme activité complexe consistant à passer de prémisses à des conclusions, plutôt qu'à une séquence causale d'intuitions, selon la définition de Pascal Engel que j'endosse tout à fait. Elle échappe ainsi au diagnostic suivant de Pascal Engel : « Ce faux rationalisme nous donne des causes des comportements humains et traite toute explication générale comme une illusion. Il refuse, explicitement ou non, la réflexivité et l'universalité de la raison, et privilégie sa face externe au détriment de sa face interne. » (p. 315) Une telle approche psychologique me semble ainsi parfaitement compatible avec les exigences de l'internalisme et susceptible de nourrir une épistémologie authentiquement attentive aux pratiques et à la situation épistémique des agents.

6 VERS UNE ÉPISTÉMOLOGIE ANCRÉE DANS LA PRATIQUE

Pour conclure, il est temps de revenir au point fondamental sur lequel je crois être en profond accord avec Pascal Engel : si l'on admet que les humains sont irrationnels et que cela est irrémédiable (ou que le remède n'est pas offert par la raison), on renonce à la raison. J'ai toutefois tenu à souligner que, si l'on veut aller jusqu'au bout de cette idée et lui donner une prise sur la réalité humaine, on doit admettre la possibilité d'une psychologie expérimentale rationaliste. Non seulement cette dernière doit être possible, mais il me semble qu'une conception épistémologique de la rationalité qui ne serait pas attentive à la pratique et aux capacités cognitives réelles des humains – une épistémologie qui énoncerait des normes de rationalité impossibles à mettre en œuvre par nos systèmes cognitifs – ne satisferait pas le « quatrième pilier du rationalisme », celui de l'accès interne aux raisons. En ce sens, la

qui, selon lui, doit opérer au moment où l'énoncé est compris et ne requiert pas d'acceptation préalable. Nos résultats parlent en faveur de l'hypothèse de Mercier.

psychologie expérimentale du raisonnement doit être une partie intégrante de l'épistémologie pour garantir une des exigences fondamentales énoncées par Pascal Engel – au nom de laquelle, précisément, il condamnait les conclusions de la psychologie.

Mais alors il faut admettre que la psychologie puisse utiliser la notion de rationalité en un sens qui satisfasse le philosophe. Or le livre de Pascal Engel est traversé par la double exigence d'une part de la garantie d'un accès interne des sujets aux raisons, qui implique selon moi une attention à la situation épistémique réelle des agents, et d'autre part d'une forme d'externalisme des raisons, destiné à nous prémunir contre le scepticisme et le pragmatisme. Cette seconde exigence semble irrémédiablement en porte à faux avec l'idée d'une enquête expérimentale sur la faculté de raison, par la notion même de raison qu'elle semble impliquer. Explorer cette tension mériterait à son tour un long développement que je laisse pour une autre occasion.

BIBLIOGRAPHIE

- Bronner, G. (2013), *La Démocratie des crédules*, Paris, PUF.
- Coady, C.A.J. (1973), « Testimony and observation », *American Philosophical Quarterly*, 10(2), 149-155.
- Cohen, L.J. (1981), « Can human irrationality be experimentally demonstrated? », *Behavioral and Brain Sciences*, 4, 317-370.
- Corner, A. et Hahn, U. (2013), « Normative theories of argumentation: Are some norms better than others? », *Synthese*, 190, 3579-610.
- Funder, D.C. (2000), « Gone with the wind: Individual differences in heuristics and biases undermine the implication of systematic irrationality », *Behavioral and Brain Sciences*, 23, p. 673-4.
- Gigerenzer, G. (2000), *Adaptive thinking. Rationality in the real world*, Oxford University Press.
- Gilbert, D.T., Krull, D.S. et Malone, P. (1990), « Unbelieving the unbelievable. Some problems in the rejection of false information », *Journal of Personal and Social Psychology*, 59(4), 601-13.
- Gilbert, D.T. (1991), « How mental systems believe », *American Psychologist*, 46(2), 107-19.
- Hahn, U. et Harris, A.J.L. (2014), « What does it mean to be biased: motivated reasoning, and rationality », *Psychology of Learning and Motivation*, 61, 41-102.
- Hahn, U. et Oaksford, M. (2006), « A normative theory of argument strength », *Informal Logic*, 26(1), 1-24.
- Harris, A.J.L. et Hahn, U. (2011), « Unrealistic optimism about future life events: a cautionary note », *Psychological Review*, 118(1), 135-54.
- Kessler, E. D., Braash, J. L. G., et Kardash, C. M. (2019), « Individual differences in revising (and maintaining) accurate and inaccurate beliefs about childhood vaccinations », *Discourse Processes: A Multidisciplinary Journal*, 56(5-6), 415-28.
- Lord, C.G., Ross, L. et Lepper, M.R. (1979), « Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence », *Journal of Personality and Social Psychology*, 37 (11), 2098-109.
- Kahneman, D. (2003), « A perspective on judgment and choice: Mapping bounded Rationality », *American Psychologist*, 58(9), 697-720.

- Kahneman, D. et Tversky, A. (1982), « On the study of statistical intuitions », *Cognition*, 11, 123-41.
- Krueger, J.I. et Funder, D.C. (2004), « Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition », *Behavioral and Brain Sciences*, 2004, 27 (3), 313-27.
- Mandelbaum, E. (2014), « Thinking is believing », *Inquiry: An interdisciplinary Journal of Philosophy*, 51(1), 55-96.
- Marr, D. (1982), *Vision: a computational investigation into the human representation and processing of visual information*, New York, W. H. Freeman and Company.
- Mercier, H. (2017), « How gullible are we? A review of the evidence from psychology and social science », *Review of General Psychology*, 21(2), 103-122.
- Mercier, H. et Sperber, D. (2017), *The Enigma of reason*, Harvard University Press.
- Oaksford, M. et Chater, N. (2007), *Bayesian rationality: the probabilistic approach to human reasoning*, Oxford, Oxford University Press.
- Oaksford, M. et Chater, N. (2020), « New paradigms in the psychology of reasoning », *Annual Review of Psychology*, 71, 305-30.
- Olsson, E.J. (2013), « A Bayesian simulation model of group deliberation and polarization », in F. Zenker (dir.), *Bayesian argumentation*, New York, Springer, 113-33.
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J. et Fugelsang, J. A. (2015), « On the reception and detection of pseudo-profound bullshit », *Judgment and Decision making*, 10(6), 549-563.
- Pinker, S. (2018), *Enlightenment now: The case for reason, science, humanism, and progress*, New York, Viking.
- Risen, J. (2016), « Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions », *Psychological Review*, 123(2), 182-207.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G., et Wilson, D. (2010), « Epistemic vigilance », *Mind & Language*, 25(4), 359-93.
- Street, C. N. H., et Kingstone, A. (2017), « Aligning Spinoza with Descartes: An informed Cartesian account of the truth bias », *British Journal of Psychology*, 108(3), 453-66.
- Tversky, A. et Kahneman, D. (1974), « Judgment under uncertainty: heuristics and biases », *Science*, 185 (4157), 1124-31.
- Vorms, M. (2021), « Bayes et les biais. Le "biais de confirmation" en question », *Revue de métaphysique et de morale*, 4(112), 567-90.
- Vorms, M., Harris, A.J.L., Topf, S. et Hahn, U. (2022), « Plausibility matters: A challenge to Gilbert's "Spinozan" account of belief formation », *Cognition*, 220.
- Weinstein, N.D. (1980), « Unrealistic optimism about future life events », *Journal of Personality and Social Psychology*, 39 (5), 806-20.