

LA NATURE DES ÉMOTIONS : QUELQUES MISES À JOUR CRITIQUES

Samuel Lepine

*Univ. Clermont Auvergne (PHIER)
samuel.lepine@uca.fr*

MOTS-CLÉS

Théorie motivationnelle des émotions; Conditions de correction des émotions; Découvertes émotionnelles; Expériences transformatrices; Raisons inarticulées; Théorie attitudinale des émotions; Théorie réactive des émotions; Objets formels de l'émotion; Représentation de valeurs; Émotions inconscientes; Justification.

1 INTRODUCTION

Il est devenu monnaie courante pour les développeurs de jeux vidéo de sortir un jeu dans un état qui n'est pas tout à fait terminé, comportant encore des *bugs* parfois plus ou moins nombreux, puis de sortir régulièrement ensuite des « *patches* » correctifs contenant des mises à jour nécessaires à l'optimisation du jeu. Si cette pratique peut paraître douteuse d'un point de vue commercial, j'ai souvent souhaité qu'il puisse exister un équivalent à celle-ci en philosophie lorsque j'écrivais *La Nature des émotions* (ci-après LNE), et plus encore une fois sa rédaction terminée. Non pas que je veuille modifier les thèses centrales, mais certains arguments m'ont paru rapidement prendre de l'âge, d'autant qu'entre la date de première rédaction et celle de publication, de nombreux articles et ouvrages sont parus, dont j'aurai aimé pouvoir tenir compte, et qui m'auraient indubitablement aidé à préciser certains des points potentiellement originaux de l'ouvrage, ou surtout à raffermir ceux qui me paraissent les plus fragiles.

Les trois contributions d'Alexis Delamare, Catherine Rioux et Artūrs Logins réunies dans ce volume ciblent manifestement des aspects du livre qui

tombent sous ces deux registres. Je les remercie d'emblée pour le temps qu'ils ont bien voulu consacrer à la lecture de ce livre, et aux critiques fines et précises qu'ils en proposent. J'espère que les *patches* correctifs que je vais proposer ici permettront d'améliorer l'ensemble, et convaincront les lecteurs qui avaient déjà consacré du temps à la lecture de l'ouvrage qu'ils n'ont pas investi dans un produit tout à fait irrécupérable (comme peuvent l'être certains jeux à leur sortie d'ailleurs)!

Quels sont les points principaux qui font l'objet des discussions ici? Le premier concerne la théorie motivationnelle que je développe quant au lien que les émotions entretiennent avec les valeurs, et qui détermine les conditions de correction des émotions. Catherine Rioux suggère qu'une approche moins subjectiviste que la mienne est possible, et insiste sur des critères de correction (notamment celui de proportionnalité) dont je ne tiens pas suffisamment compte. Le deuxième point concerne la théorie des émotions comme attitude évaluatives, que j'emprunte à Julien Deonna et Fabrice Teroni (2012). Celle-ci est devenue de plus en plus une théorie centrale en philosophie des émotions, et a fait par conséquent l'objet de discussions et critiques abondantes. La lecture de LNE que propose Alexis Delamare s'en fait l'écho, puisqu'il suggère d'abandonner le modèle attitudinaliste pour un modèle « réactif », où les émotions ne sont plus des évaluations, mais des réactions à des évaluations qui les précèdent. Enfin, Artūrs Logins discute à la fois la théorie de la justification des émotions que je propose, mais surtout du cas des émotions inconscientes. Il soutient, en reprenant un argument « anti-luministe » de Williamson, qu'il est possible qu'il y ait des émotions inconscientes.

En répondant à ces trois discussions de mon travail, j'espère non seulement pouvoir « patcher » certaines des « failles critiques » de LNE, mais surtout montrer que les thèses qui y sont élaborées peuvent prendre une tournure que je n'ai encore pas développée jusqu'ici. J'esquisserai donc, dans mes réponses, certaines des directions futures que pourraient prendre ces thèses. Je commence, dans la première section, par répondre à Catherine Rioux, en revenant sur les questions qui concernent la partie la plus originale de LNE, à savoir la défense d'une théorie motivationnelle des émotions et ses implications, et j'en profite pour distinguer deux types de théories motivationnelles. Je réponds ensuite à Alexis Delamare en dressant un état de l'art des débats entourant la théorie attitudinale des émotions, et des développements récents qui ont entouré cette dernière et qui me poussent à continuer de penser que cette théorie reste la meilleure option. J'aborde enfin les difficultés que soulève Artūrs Logins relativement à la justification des émotions dans le cadre de la théorie motivationnelle, ainsi que la question des émotions inconscientes.

J'espère, dans tous les cas, que mes réponses seront à la hauteur de la qualité des lectures qui ont été proposées de LNE, et qu'elles pourront donner lieu encore à de plus amples échanges à l'avenir. Je remercie enfin Vincent Boyer et Sébastien Motta pour avoir eu la patience et la bienveillance d'organiser une journée d'étude et ce volume autour de LNE.

2 COMMENT CONCEVOIR LA THÉORIE MOTIVATIONNELLE ? RÉPONSE À CATHERINE RIOUX

Catherine Rioux concentre sa discussion sur l'une des thèses centrales de LNE – peut-être la plus originale à mon sens – à savoir la défense de la théorie motivationnelle, qu'elle se propose d'endosser également mais sous une forme un peu différente ou allégée. Rioux remarque notamment que cette théorie n'est pas nécessairement liée à une option métaphysique particulière au sujet de l'existence des valeurs. De même que l'approche indépendantiste (selon laquelle les conditions de correction d'une émotion sont indépendantes de notre structure motivationnelle) ne va pas nécessairement de pair avec le réalisme au sujet des valeurs, l'approche motivationnelle (selon laquelle les conditions de correction d'une émotion dépendent de notre structure motivationnelle) ne va pas nécessairement de pair avec une approche antiréaliste ou subjectiviste au sujet des valeurs. Rioux, au contraire, soutient que l'approche motivationnelle est compatible avec une forme de réalisme. Je suis assez largement d'accord avec cette thèse, et j'y reviendrai un peu avant la conclusion. Je ne crois donc pas qu'il y ait de véritable désaccord ici.

Les points de désaccords concernent plutôt deux autres aspects de la théorie motivationnelle. L'un porte sur la distinction entre des conditions de correction simples et complexes que j'esquisse dans LNE, et dont Rioux conteste la clarté et la pertinence ; et l'autre porte sur les éléments à prendre en considération dans la correction des émotions dans le cadre d'une approche motivationnelle : là où je propose de tenir compte des états motivationnels qui sont antérieurs à l'émotion (comme le désir, les sentiments, les préférences, etc.), Rioux insiste sur les aspects motivationnels qui sont constitutifs de l'émotion, comme les tendances à l'action qui vont avec cette dernière. Cette discussion nous permettra ainsi de distinguer deux approches motivationnelles possibles de la correction, qui ne sont cependant pas nécessairement exclusives. Je montre en tout cas que certains aspects de la théorie de Mason – que mobilise Rioux – peuvent être intégrés dans la théorie motivationnelle de LNE, quoique je rejette *in fine* sa dimension représentationnelle. Je reviens enfin rapidement sur la question des engagements métaphysiques qui accompagnent mon approche motivationnelle.

2.1 Y A-T-IL DES CONDITIONS DE CORRECTION « SIMPLES » ?

Un premier véritable point de discordance concerne la question des justifications que je propose en faveur de l'approche motivationnelle, et notamment le fait que je distingue des conditions de correction simples et complexes. En effet, je propose cette distinction pour rendre compte des cas où certaines émotions semblent donner du poids à l'approche indépendantiste, comme lorsque l'on s'émerveille d'une mélodie pour laquelle nous n'avons pas d'intérêt préalable ou que nous sommes touchés par l'attitude d'une personne contre laquelle nous éprouvons par ailleurs de l'hostilité. Selon l'approche

indépendantiste, ces découvertes émotionnelles nous révéleraient ainsi certaines valeurs indépendamment de notre structure motivationnelle, et cela semble être une vertu de cette approche que de pouvoir rendre compte de façon intuitive de ce genre de cas assez courants.

Ma réponse, dans LNE, était que cette impression de découverte émotionnelle était liée à deux choses. Premièrement au fait que ces expériences émotionnelles s'appuient en réalité sur certaines motivations « basiques » que nous partageons tous, et qui confèrent une certaine évidence à ces découvertes, qu'il est plus difficile de retrouver dans les motivations moins basiques, liées à notre culture ou notre biographie personnelle. Ces motivations seraient basiques au même sens qu'il y a des émotions dites « basiques »¹ : elles sont distribuées de façon relativement uniforme et transculturelle au sein de l'espèce humaine, et peuvent s'expliquer par notre passé évolutif. Il est probable que nous ayons une motivation basique à éviter le danger ou les sources de contamination, une autre à souder des liens avec nos proches pour créer des relations de coopération et de reproduction, une motivation à rechercher et apprécier la nourriture, et probablement aussi certaines motivations esthétiques.

Cette liste n'est pas exhaustive, elle est au contraire ouverte, et compatible avec une théorie évolutionnaire au sujet des motivations constitutives de la nature humaine. Le fait est que ces motivations basiques engendrent – et c'est le deuxième point – des conditions de corrections « simples », au sens où elles sont peu sujettes à discussion, et laissent moins de place aux disputes évaluatives, précisément parce qu'elles reposent sur des motivations que nous avons tous en partage. S'il nous semble évident que certaines émotions sont correctes, et qu'elles nous « révèlent » des valeurs indépendamment de nos motivations, c'est parce que ces valeurs sont au contraire liées à des motivations très fortement enracinées en nous. Ce que ces « découvertes émotionnelles » nous révèlent alors, c'est bien plutôt la présence de ces motivations basiques en nous.

Rioux exprime un certain scepticisme en regard de cette discussion : où se situe en effet le seuil entre des conditions de correction simples et complexes, et quel est le critère qui permet de les distinguer ? Par ailleurs, ne peut-il pas y avoir des découvertes émotionnelles liées à des conditions de correction complexes ? Sur ce dernier point, les formules de LNE étaient probablement sinueuses, et il me semble donc bienvenu de reformuler ici les choses. Commençons par le premier point. La notion de conditions de correction « simples » renvoie à l'idée que certaines émotions dépendent de mécanismes cognitifs de bas-niveau, auxquelles nous sommes disposés de façon innée en regard de certains *stimuli* que l'on peut retrouver dans certains produits esthétiques mais aussi dans des paysages, des interactions humaines, etc.

1. Voir LNE, chapitre III.

Le fait que ces émotions soient fortement liées à notre équipement évolutif, pour ainsi dire, et liées à des *stimuli* relativement simples d'un point de vue cognitif (certaines harmonies de couleurs, de sons, des rapports humains chaleureux, etc.), contribue à faire en sorte qu'on peut facilement se mettre d'accord sur leur caractère approprié. Le fait que certaines œuvres viennent chercher ces émotions de bas-niveau me semble d'ailleurs attesté par le fait qu'elles sont souvent l'objet de critiques esthétiques estimant que les artistes ne devraient pas s'abaisser à éveiller des émotions aussi simples avec des musiques, des romans, ou des films sirupeux, larmoyants, niais, plein de bons sentiments, etc².

Les choses semblent moins évidentes avec les émotions dépendant de processus cognitifs de plus haut-niveau comme celles qui sont liées à nos croyances au sujet de ce qui constitue un bon polar, un jeu d'aventure réussi, une exploitation intelligente des règles dans un sport, ou encore une action moralement bonne d'un point de vue normatif élaboré, et qui dépendent généralement de motivations qui ont un arrière-plan plus fortement culturel ou personnel qu'inné. Dans ce genre de situations, les considérations susceptibles de compter en faveur de la correction d'une émotion sont plus nombreuses et élaborées, et c'est en ce sens que leurs conditions de corrections sont plus « complexes ». Mais il est clair qu'il n'y a pas pour autant de seuil précis entre les conditions de correction simples et complexes, et que la différence entre les deux est une différence de degré plutôt qu'une différence de nature. Disons simplement que les conditions de correction semblent plus simples si elles renvoient à des motivations basiques et que la correction de l'émotion ne semble que peu sujette à discussion. Elles deviennent de plus en plus complexes lorsqu'elles dépendent de nos croyances et surtout de considérations normatives à l'égard de ce que doit être une œuvre d'art, un beau geste sportif, une action moralement acceptable, ou encore une vie accomplie.

Ceci n'exclut cependant pas qu'il puisse y avoir des découvertes émotionnelles avec ces émotions plus complexes, comme je le reconnaissais dans LNE. Mais celles-ci supposent un arrière-plan axiologique différent. Tandis que les émotions relatives aux motivations basiques ne présupposent pas de notre part de croyances, les découvertes émotionnelles plus complexes prennent forme sur un arrière-plan de croyances : pour s'émerveiller de l'originalité d'une œuvre, d'une performance artistique ou sportive, il faut un certain nombre de croyances à leur égard. On ne peut pas être saisi d'admiration envers Steve Nash et découvrir émotionnellement la valeur de son

2. Ce n'est bien sûr pas la seule raison, et d'autres critiques portent moins sur le caractère « populaire » de certaines de ces œuvres, que sur le fait qu'elles véhiculent des valeurs implicites discutables, avec par exemple une vision manichéenne du monde, une conception de l'amour simpliste, etc. Il n'est pas rare que les deux aspects soient liés, et que l'on reproche à des œuvres de véhiculer ces valeurs « sous le manteau », pour ainsi dire, en faisant appel à nos motivations basiques, mais il n'y a aucune raison de penser que l'exploitation de ces dernières implique nécessairement ce point plus discutabile.

exploit en apprenant qu'il a réalisé quatre saisons en 50-40-90 (soit 50% de réussite au tir, 40% à trois points, et 90% aux lancers-francs), sans savoir ce qu'implique une telle statistique d'un point de vue historique. De même, tous les triples sauts se ressemblent probablement pour un non-initié, qui aura du mal à percevoir ainsi le caractère incroyable des performances de Jonathan Edwards. Enfin, il est parfois difficile d'apprécier certaines œuvres sans connaître leur histoire et sans s'intéresser à celle-ci. Elles ont pu être avant-gardistes à une époque et mal vieillir ensuite, de sorte qu'il faut un effort cognitif pour découvrir leur valeur. Enfin, comme dans l'exemple que prend Rioux, on peut aussi avoir certaines croyances au sujet d'une activité qui seront prises en défaut par une émotion. L'idée n'est donc pas d'affirmer qu'il n'y a pas de découvertes émotionnelles de haut-niveau, mais que celles-ci font probablement moins l'objet d'un consensus que les découvertes de bas-niveau, et que si l'on tient compte de ce fait, alors il est peut-être moins plausible que les conditions de corrections de nos émotions soient réellement indépendantes de nos motivations.

2.2 LES DÉCOUVERTES ÉMOTIONNELLES TRANSFORMATRICES ET LES RAISONS INARTICULÉES

Une autre objection possible sur laquelle je passe peut-être trop rapidement dans LNE, à cet égard, est le fait que nos découvertes émotionnelles semblent souvent entraîner de nouvelles motivations, là où je soutiens au contraire dans LNE que nos découvertes émotionnelles nous révèlent des motivations qui étaient déjà là. Bien que Rioux ne mentionne pas cette difficulté, il me semble important de revenir dessus dans le cadre de cette discussion, car je crois que c'est un prolongement de la discussion qui précède, et sur lequel on pourrait être tenté de s'appuyer pour défendre la thèse indépendantiste contre l'approche motivationnelle.

En effet, ces découvertes émotionnelles qui ne s'intègrent pas dans nos états motivationnels préexistants donnent parfois lieu à des « expériences transformatrices » (Paul, 2014, ou 2023 pour la traduction), au sens où elles modifient la structure de nos préférences d'une façon qui n'était pas rationnellement prévisible. Pour reprendre l'exemple classique de Paul (2014), avoir un enfant est susceptible de changer à la fois la personne que je suis (incluant mes préférences, la manière dont je souhaite désormais dépenser mon temps, etc.), ainsi que ma position épistémique : je découvre quelque chose en faisant l'expérience de la parentalité que je ne pouvais pas savoir avant, puisque ce savoir est avant tout phénoménal, tout comme c'est le cas lorsque l'on tombe amoureux par exemple³.

3. J'avais essayé d'aborder ce point dans Lepine (2016) à une époque où je ne connaissais pas encore l'ouvrage de Paul (2014), en prenant précisément l'exemple de la parentalité.

Or, intuitivement, ce genre de cas semble plaider en faveur de l'approche indépendantiste : si ces expériences sont transformatrices, n'est-ce pas parce qu'elles nous révèlent la valeur de certaines expériences indépendamment de nos motivations ? Il me semble au contraire que ce serait une simplification abusive de ce que sont ces expériences émotionnelles transformatrices. Il est vrai cependant que ces expériences ne nous révèlent pas non plus des motivations qui seraient déjà là, comme le voudrait une application naïve de la théorie motivationnelle, puisqu'il est clair que je n'étais pas littéralement enclin à m'épanouir comme parent avant d'en faire l'expérience. Elles sont transformatrices précisément parce qu'elles modifient en profondeur nos motivations, et créent ainsi des changements qui ne s'intègrent pas immédiatement dans notre structure motivationnelle. Elles nous mettent ainsi en présence de ce que j'appelle ailleurs dans LNE des « raisons inarticulées » (p. 221), et ce concept peut nous permettre de comprendre pourquoi l'approche motivationnelle n'est pas en si mauvaise posture face à ce genre de cas.

Ces raisons inarticulées ne sont pas en mesure de justifier nos émotions à elles seules, mais peuvent cependant jouer par rapport aux découvertes émotionnelles transformatrices un rôle analogue à ce que Agnès Callard (2018) nomme des « raisons proleptiques ». Les raisons proleptiques, telles que les définit Callard sont en effet des raisons provisoires, qui manifestent une compréhension inchoative d'un phénomène que l'on va chercher à mieux connaître. Elles seraient emblématiques, selon Callard, des situations dans lesquelles nous aspirons à quelque chose sans avoir une représentation claire de ce à quoi nous aspirons (s'engager dans la poursuite d'une activité telle que la philosophie, ou dans une relation dont nous devinons à peine les conséquences réelles et les implications pratiques), alors que la poursuite de cette chose est susceptible de transformer en profondeur la personne que nous sommes et les choses auxquelles nous tenons. J'ai une raison proleptique de m'engager dans des études de philosophie lorsque je me base sur ce que j'espère devenir grâce à ces études, alors même que je n'ai qu'une représentation un peu vague de la valeur liée à ces études. Je ne peux donc pas vraiment justifier mon action sur la base de ce que cela m'apportera, mais seulement sur l'espoir que j'ai d'obtenir ce bien et d'en retirer certains bienfaits. Le but visé n'est pas ici véritablement connu à l'avance, et c'est à mesure que l'on suit cette « aspiration » que l'on espère en apprendre davantage et éclaircir les raisons que l'on a de la poursuivre.

Mais c'est précisément sur ce point que les raisons inarticulées se distinguent des raisons proleptiques, qui n'en sont d'une certaine façon qu'une sous-catégorie. Les raisons proleptiques font en effet appel à des motivations sur le long-terme (la personne que je veux devenir) et ont en ce sens une dimension existentielle, là où les raisons inarticulées font appel à des motivations qui n'ont pas nécessairement un lien décisif avec la personne que nous aspirons à être, mais simplement avec des expériences dont nous pourrions apprendre à apprécier la valeur. Elles possèdent certes, comme les raisons proleptiques, une dimension exploratoire, dans la mesure où elles nous

poussent à rechercher une expérience non pas en vue d'un bien que l'on désire, mais pour en apprendre plus sur notre propre relation à cette expérience. Mais elles visent avant tout la congruence possible d'une motivation naissante avec d'autres motivations que nous avons déjà. Dans la parentalité, je peux par exemple renouer avec certains plaisirs anciens (la joie de construire des choses avec des petites briques en plastique), apprécier le fait de s'émerveiller de choses qui étaient devenues pour nous anecdotiques, ou encore me réjouir de partager des œuvres que je n'avais plus vues ou écoutées depuis longtemps. C'est ainsi que l'expérience de la parentalité, et le désir subséquent de s'engager dans cette expérience, viennent s'insérer dans une structure motivationnelle qui leur préexistait.

En ce sens, ces expériences transformatrices ne nous révèlent pas non plus une valeur indépendante de notre structure motivationnelle : elles nous mettent en présence de raisons inarticulées, qui pourront venir s'articuler avec certaines de nos motivations, mais qui dans d'autres circonstances pourraient ne pas s'articuler à celles-ci. Il se pourrait par exemple que j'aie des motivations contraires trop fortes, qui pourraient me rendre insensible à la force de ces raisons et les rendre plus opaques encore. Aveuglé par mon attachement à ma carrière, je pourrais rester insensible aux joies de la parentalité. De façon moins tragique (et moins biaisée peut-être), il arrive fréquemment que des expériences plaisantes ne donnent pas lieu à une transformation de soi, parce que nous n'avons pas une « place » suffisante à accorder à de nouvelles motivations dans nos vies. De même, je peux découvrir qu'une certaine expérience est agréable, et y rester partiellement indifférent, par exemple en découvrant un type d'art que je ne connaissais pas. Il y a donc découverte émotionnelle au sens où je le propose dans LNE, dans la mesure où dans ce genre de situations, je découvre qu'une expérience donnée pourrait s'accorder avec certains de mes intérêts ou de mes motivations, et peut même me révéler par la même occasion que j'ai ces intérêts ou motivations. Si de telles expériences ne donnent toutefois pas nécessairement lieu à une transformation de soi, c'est probablement parce que l'on ne se reconnaît pas soi-même suffisamment dans les motivations qu'elles révèlent, autrement dit parce que ces motivations elles-mêmes ne sont pas suffisamment fortes en nous pour qu'on leur accorde davantage de place ou d'importance.

Mais les expériences émotionnelles transformatrices ne nous révèlent pas non plus à proprement parler des motivations préexistantes : elles nous révèlent plutôt comment certaines motivations préexistantes peuvent s'articuler avec de nouvelles expériences pour donner naissance à de nouvelles motivations. Ainsi, ma passion pour les choses difficiles peut me pousser à aimer les jeux qui demandent beaucoup d'entraînement ou de concentration, mais peut aussi trouver de nouvelles satisfactions dans des disciplines sportives exigeantes. Là où les raisons proleptiques ont donc une dimension motivationnelle sur le long terme, les raisons inarticulées ont une dimension beaucoup plus court-termiste. J'ai une raison proleptique de poursuivre une activité, soutient Callard, du fait que j'entrevois de façon confuse la valeur que

pourrait avoir cette activité pour moi. Les raisons inarticulées n'ont pas cette dimension d'aspiration, et nous invitent plus prosaïquement à introduire de la cohérence dans notre structure motivationnelle.

En ce sens, les émotions liées aux expériences transformatrices ont probablement tendance à fonctionner selon le modèle proposé autrefois par Oatley et Johnson-Laird (2002). Selon eux, les émotions engendreraient en effet des raisonnements abductifs destinés à réviser et ajuster nos croyances et motivations lorsque nous détectons de l'incohérence entre celles-ci. D'une certaine façon c'est une prédiction plausible y compris pour la théorie psychologique de l'*appraisal* (que je défends dans LNE) qui stipule que les émotions apparaissent avec la détection d'un événement pertinent pour nos motivations. Si cet événement n'est pas congruent avec nos motivations et/ou nos croyances, alors il nous conduira dans certains cas à réviser celles-ci. C'est typiquement ce qui se passe dans certains des cas décrits plus haut. Mais notons que l'exemple inverse est possible : l'étudiant de philosophie qui aspirait à devenir un grand penseur, déjà prêt à écrire la prochaine somme totalisante au sortir des classes préparatoires, pourrait éprouver une amère déception en découvrant que la philosophie contemporaine ne correspond pas à ce qu'il avait imaginé. La déception de notre étudiant lui donnerait alors une raison inarticulée d'abandonner son projet. Mais peut-être échouera-t-il à intégrer cette raison dans sa structure cognitive et motivationnelle parce que son amour pour la philosophie le poussera à persévérer et à découvrir des plaisirs insoupçonnés dans la philosophie contemporaine.

Une autre manière de comprendre ce que sont ces raisons inarticulées serait alors de les indexer sur le concept de « raison *pro tempore* » que propose Brady (2013). Selon lui, toutes les émotions nous donneraient en effet des raisons de ce type, qui sont des raisons provisoires, qui demandent à être complétées par une réflexion critique plus générale, et qui seulement alors pourront contribuer à notre connaissance évaluative. Cela revient à dire, d'une certaine façon, que les émotions ne joueraient un rôle que secondaire dans cette connaissance, puisque leur contribution nécessiterait le filtre de notre regard critique⁴. En un sens, on pourrait être tenté de penser que les découvertes émotionnelles, telles que je les ai présentées ici, ne nous donnent que des raisons provisoires de ce type, et que les raisons inarticulées ne seraient qu'un autre nom pour ces raisons *pro tempore*. La thèse de Brady est d'ailleurs présentée par Silva (2021b) comme une thèse « motivationnelle » au sujet de la justification, puisque les émotions, dans ce cadre, nous motivent à chercher des raisons qui pourront les justifier (ou au contraire leur retirer leur justification), mais ne nous donnent pas immédiatement de justification. Silva oppose cette conception à la thèse plus standard selon laquelle les émo-

4. Pour une présentation critique de l'analyse de Brady, je me permets de renvoyer à Lepine (2022).

tions pourraient justifier nos jugements de valeur dans certaines conditions (une thèse que je défends partiellement dans LNE).

Mais les raisons inarticulées ne sont pas des raisons provisoires en ce sens-là non plus. Après tout, les découvertes émotionnelles nous donnent des raisons de nous engager dans certaines activités qui ne nécessitent pas nécessairement de vérification sur leur bien-fondé, comme le recommande Brady au sujet des raisons *pro tempore*. Elles sont inarticulées simplement au sens où elles ne sont pas immédiatement articulées avec notre structure motivationnelle, et nécessitent que nous puissions les articuler avec celle-ci pour pouvoir nous engager davantage dans l'activité qu'elles motivent. Ce sont donc des raisons inarticulées en un sens pratique, mais aussi en un sens épistémique : si elles sont en effet justifiées du point de vue de leur objet (l'expérience nouvelle x me donne de bonnes raisons d'éprouver de la joie), leur justification reste indécise du point de vue de nos motivations tant qu'aucun état motivationnel ne nous incite à les valoriser. C'est cette discordance possible, dans les découvertes émotionnelles, qui fait que ces raisons inarticulées nous motivent en même temps à réviser nos croyances ou motivations.

Ainsi, pour terminer sur ce point, les expériences émotionnelles transformatrices ne me semblent pas nécessairement plaider plus en faveur de la thèse indépendantiste que de la thèse motivationnelle, puisque cette dernière peut au contraire expliquer assez simplement pourquoi certaines découvertes émotionnelles ne sont pas transformatrices.

2.3 DEUX VERSIONS DE LA THÈSE MOTIVATIONNELLE

Venons-en au deuxième volet de la lecture critique que Rioux propose de LNE. Une autre manière de défendre la thèse motivationnelle, selon elle, serait d'insister sur une « contrainte de proportionnalité » dans la correction de l'émotion, comme le défend Mason (2025). Selon Mason, en effet, il ne suffit pas qu'une émotion représente adéquatement son objet pour être appropriée ou correcte, mais il faut encore qu'elle soit proportionnelle aux motivations qu'elle implique et que ces motivations soient correctes. Certains auteurs, comme Tappolet (2016) ont déjà proposé de tenir compte de cette contrainte de proportionnalité en termes représentationnels : une peur correcte serait par exemple une peur dont l'intensité est corrélée à l'intensité du danger perçu. Une peur intense serait ainsi incorrecte si elle représente un danger mineur. Mais l'idée de Mason est que cet élément représentationnel est insuffisant pour rendre compte de la proportionnalité, puisque l'intensité de notre motivation émotionnelle devrait aussi être proportionnelle au danger : si le danger est très élevé, alors je devrais aussi être fortement motivé à l'éviter pour que ma peur soit correcte.

Un point auquel il faut être attentif d'emblée est que le concept de motivation tel que Mason l'emploie renvoie aux motivations comportementales qui accompagnent l'émotion, et qui se traduisent par une « préséance de contrôle » (*control precedence*) dans l'orientation de notre attention et de nos tendances

à l'action, ainsi que dans le traitement de l'information : fuir dans le cas de la peur, réparer sa faute pour la culpabilité, exhiber l'objet de notre réussite pour la fierté, etc. Ces différents éléments participent tous à la « priorisation » de la motivation à laquelle ils appartiennent, et qui fait partie de l'émotion. Une émotion correcte, selon cette approche, est donc une émotion qui non seulement représente correctement la valeur de son objet, mais qui manifeste une motivation appropriée à l'égard de cette valeur. Être obsédé par l'objet d'une fierté pourtant mineure, au point de vouloir en parler à tout le monde, serait en ce sens inapproprié, de même qu'il serait inapproprié d'être hanté par les remords à l'égard d'une faute que tout le monde s'accorde à reconnaître comme sans importance. Dans les deux cas, il n'y a pas de bonnes raisons qui plaident en faveur de cet aspect motivationnel de l'émotion.

Mason propose donc une théorie « hybride » où une émotion est correcte lorsqu'elle représente fidèlement la valeur de son objet, et que ses aspects motivationnels sont soutenus par des raisons liées à la proportionnalité de cette émotion. Cette théorie a ainsi le mérite de tenir compte des tendances comportementales qui accompagnent l'émotion dans sa caractérisation de la correction. Mais il est clair que les motivations dont il est question ici ne sont pas les mêmes que celles qui sont au cœur de la théorie motivationnelle de LNE, puisqu'elles renvoient à ce qui se produit en aval de l'émotion, là où la théorie motivationnelle s'intéresse aux états motivationnels qui interviennent en amont, tels que les désirs, les sentiments, les préférences, etc. En réalité, la théorie de Mason s'inscrit dans la droite lignée d'une autre « théorie motivationnelle » au sujet des émotions, défendue par Scarantino (2014). Scarantino, en effet, défendait déjà l'idée qu'une théorie satisfaisante des émotions devrait prendre en compte leur aspect motivationnel du point de vue de la préséance qu'elles exercent sur notre comportement, et ne pas se focaliser uniquement sur leur rapport aux valeurs.

La théorie de Mason, telle que je la comprends, est une extension de la théorie motivationnelle de Scarantino à la question des conditions de correction de l'émotion. Un composant motivationnel comme l'attention, selon Mason, peut être correct, pour autant qu'il soit ajusté à la valeur de l'objet de l'émotion. Cette théorie n'est bien sûr pas incompatible avec la prise en compte des états motivationnels dans la correction des émotions, mais ce n'est pas sa revendication première (et il ne semble pas que ce soit une revendication explicite de Mason, qui emploie des formules généralement proches de l'indépendantisme). Inversement, la théorie motivationnelle défendue dans LNE s'efforce de préciser la nature des conditions de correction des émotions en s'intéressant à leur étiologie motivationnelle. Du point de vue de cette théorie, une émotion peut être considérée comme correcte lorsqu'elle est justifiée par les propriétés de son objet, et qu'elle est liée à un état motivationnel qui est lui-même justifié. Par ailleurs, contrairement à la théorie hybride de Mason, je ne crois pas qu'il y ait une dimension représentationnelle dans la *fittingness*, puisque je rejette l'idée que les émotions soient des représentations de valeurs. Ce sont les émotions en tant qu'attitudes qui sont susceptibles

d'être appropriées à leur objets, et non en tant que représentations. Mais ce point reste d'une importance mineure ici.

Rioux propose néanmoins de fusionner ces deux approches motivationnelles : sa suggestion est que l'on peut défendre une théorie motivationnelle au sujet de la correction en suivant le modèle de Mason, tout en la mariant avec une approche motivationnelle telle que celle défendue dans LNE, au sens où la correction de l'espoir, pour reprendre son exemple, dépend notamment de nos désirs. Selon Rioux, cela n'exclut pas que d'un point de vue ontologique, les valeurs pourraient exister indépendamment de nous par ailleurs. Autrement dit, on pourrait à la fois défendre une thèse motivationnelle forte, et ne pas s'embarrasser des implications subjectivistes – voire potentiellement antiréalistes – qui semblent accompagner l'approche défendue dans LNE.

D'une certaine façon, je suis loin d'être opposé à ce type d'approche. La question qui m'intéresse avec la théorie motivationnelle est la question épistémologique des conditions de correction, et j'essaie de me tenir à l'écart le plus possible de considérations ontologiques au sujet des valeurs. Comme je le concède à la fin du chapitre V, cette approche pourrait très bien se marier avec certaines formes de réalisme en métaéthique, proches d'un « réalisme interne » à la Putnam, où les vérités métaéthiques seraient dépendantes de considérations élaborées sur la base de motivations communes, ainsi qu'avec certaines formes de constructivisme. Il est peu probable, en revanche, que cette approche soit compatible avec une forme de réalisme robuste où les vérités morales reposeraient sur des faits non-naturels, et indépendants de nous. De ce point de vue, même si la théorie motivationnelle n'est pas tout à fait neutre, je maintiens qu'elle est relativement malléable d'un point de vue métaphysique.

2.4 CONCLUSION

S'il me semble important de maintenir la distinction entre conditions de correction simple et complexe, je reconnais qu'elle ne repose pas sur une frontière claire, et il est possible qu'en poussant plus loin la discussion sur ce point, de nouvelles difficultés apparaissent pour la thèse motivationnelle. J'ai essayé de montrer au passage que les découvertes émotionnelles étaient susceptibles de soulever d'autres problèmes encore pour cette dernière, et notamment du point de vue des expériences transformatrices, tout en suggérant certaines réponses que pourrait envisager un partisan de l'approche motivationnelle. Enfin, les suggestions de Catherine Rioux quant à de possibles amendements de la thèse motivationnelle elle-même me semblent bienvenues, pour autant que cette thèse conserve une forme non représentationnelle. Sur ce point, les mises à jour que suggère Rioux pour la thèse motivationnelle me paraissent donc bienvenues.

3 THÉORIE ATTITUDINALE OU THÉORIE RÉACTIVE ? RÉPONSE À ALEXIS DELAMARE

Dans la discussion qu'il propose des thèses de LNE, Alexis Delamare suggère ce que l'on pourrait appeler un correctif mineur et un correctif majeur. Le correctif mineur concerne l'objet formel de la colère (c'est-à-dire la valeur visée par la colère), que je présente comme étant l'offense. Il souligne, à juste titre, que dans de nombreux cas la colère ne semble pas porter sur une offense, mais sur des situations où notre désir est frustré ou empêché. À vrai dire, je suis plutôt d'accord avec ce point, et j'ai clairement eu une hésitation à cet égard tout au long de la rédaction de l'ouvrage. D'un côté, il faut le reconnaître, l'idée que la colère porterait sur des offenses est vraiment une thèse standard : c'est une affirmation que l'on retrouve couramment sous la plume de nombreux spécialistes de la colère (voir notamment Silva, 2021a). D'un autre côté, il semble évident que l'on peut être en colère de façon justifiée sans qu'aucune offense n'ait été commise, par exemple parce que l'on ne parvient pas à construire un meuble qui était pourtant censé être simple à assembler (du moins selon la notice). Une hypothèse plausible est qu'il y ait en réalité plusieurs formes de colère, l'une relevant plus de la frustration face à l'empêchement d'un désir, l'autre étant davantage interpersonnelle et morale, et liée à des situations d'irrespect. Dans tous les cas, ce premier correctif peut être appliqué sans hésitation, même s'il ouvre en réalité sur une discussion plus large sur la nature de la colère.

Venons-en maintenant au correctif majeur que propose Delamare, et qui concerne la théorie motivationnelle élaborée dans LNE. Delamare prend au sérieux cette théorie, et souligne le point suivant : s'il est vrai, comme le veut la théorie motivationnelle, que les états motivationnels fournissent des évaluations précalibrées pour les émotions (c'est parce que j'aime ma femme que j'évalue sa victoire au tennis comme réjouissante), alors, affirme Delamare, les émotions n'ont pas de rôle évaluatif à proprement parler. Les émotions seraient plutôt des réponses à ces valeurs que nous appréhendons par le biais de nos motivations. Selon cette suggestion, ma colère serait une réponse à une situation que j'appréhende comme faisant obstruction à mon désir. En ce sens, l'objet formel de l'émotion serait déjà appréhendé avant que l'émotion n'entre en jeu, et celle-ci serait bien plutôt une réponse que l'on adopte en regard de celui-ci.

Cette suggestion permettrait aussi d'affranchir plus nettement la théorie attitudinale de ce que Delamare propose de nommer le modèle de la « cognition axiologique », dans lequel les émotions sont conçues comme des appréhensions de valeurs, comme c'est le cas typiquement pour la théorie perceptuelle des émotions. Selon cette approche, que je critique également dans LNE, les émotions nous révéleraient des valeurs, du moins lorsqu'elles sont correctes ou appropriées à leur objet, de sorte qu'il suffirait d'éprouver une peur appropriée pour être conscient d'un danger par exemple. Contre ce modèle, Delamare rappelle deux problèmes qui sont régulièrement mentionnés

dans la littérature : à savoir d'une part le fait que les émotions sont sujettes à des « Pourquoi ? », ce qui n'est pas le cas des états perceptifs ordinaires. On peut me demander pourquoi j'éprouve de la peur, et je devrais alors justifier mon émotion en faisant appel aux propriétés de l'objet. Alors qu'il n'y a aucun sens à demander pourquoi je vois un chat noir : le fait est qu'il est devant moi et que ma perception me délivre ce contenu de façon transparente. Et d'autre part, le fait qu'il soit possible d'avoir accès aux valeurs d'une façon non émotionnelle : je peux percevoir que le tableau devant moi est beau sans m'en émouvoir, ou savoir que les tigres sont dangereux sans éprouver de peur en leur présence.

À ce modèle, Delamare oppose alors le modèle de la « réponse axiologique », selon lequel l'émotion serait une réponse à un objet formel perçu par l'entremise d'un « sentiment axiologique ». L'émotion ne serait donc plus ici un genre de connaissance axiologique, mais une réponse à une valeur déjà perçue par une appréhension affective qui précède l'émotion, et qui est opérée par le sentiment (qu'il faut entendre ici au sens de *feeling*). C'est là le cœur de ce que l'on appelle communément aujourd'hui la « théorie réactive » des émotions. Selon Delamare, cette théorie serait en réalité celle qui s'accorde le mieux avec le portrait des émotions que j'esquisse dans LNE. Dans le cadre de la théorie réactive, en effet, la phénoménologie des émotions serait précisément ce qui ferait leur spécificité en tant que « réponses » à des objets formels appréhendés par nos sentiments de valeurs : les émotions seraient des réponses caractérisées par un certain ressenti spécifique, qui auraient ainsi une dimension « chaude » les distinguant du même coup des connaissances axiologiques froides. Si le concept de sentiment de valeur est souvent décrié comme mystérieux, Delamare souligne que l'approche motivationnelle proposée dans LNE permettrait précisément de lever ce mystère en suggérant que les sentiments de valeurs s'expliquent par nos motivations affectives⁵.

La proposition d'Alexis Delamare est manifestement séduisante, et la théorie réactive s'est clairement imposée au cours des dernières années comme une concurrente sérieuse de la théorie attitudinale, notamment sous la plume de Müller (2017) et de Massin (2023), qui la présentent d'ailleurs comme une variante de la théorie attitudinale. L'absence de discussion de cette théorie dans LNE pourrait ainsi légitimement sonner comme un silence coupable. Je pense toutefois que la théorie attitudinale continue de constituer une voie plus crédible que la théorie réactive, et qu'elle s'accorde par ailleurs mieux avec l'approche motivationnelle que je défends. Ceci étant, les théories de la cognition axiologique ne sont pas en reste, et la théorie perceptuelle a reçu aussi une défense vigoureuse sous la plume de Rossi et Tappolet⁶. Dans ce qui suit, j'essaie de montrer pourquoi l'approche attitudinale me semble pré-

5. Müller souligne également cette possibilité (2017, 293).

6. Voir aussi l'ouvrage de Tappolet (2016).

férable à la théorie réactive, et comment elle peut répondre aux arguments récents avancés en faveur de l'approche perceptuelle.

3.1 LES ARGUMENTS EN FAVEUR DE LA THÉORIE RÉACTIVE

La théorie réactive, on l'a vu, prend au sérieux l'idée que les émotions sont sensibles aux raisons, ou qu'elles répondent à des raisons (à des « pourquoi ? »). Mais de quelle nature sont les raisons d'éprouver des émotions ? L'approche attitudinale défendue par Deonna et Teroni, et que je reprends dans LNE, affirme que les raisons d'éprouver des émotions sont fournies par les objets de nos émotions, et sont en ce sens des raisons non-évaluatives telles que les crocs du chien (pour la peur), la mort d'un proche (pour la tristesse), ou mon comportement après trois verres (pour la honte). Et c'est en tant que ces propriétés non-évaluatives sont susceptibles de constituer un danger – ou d'être les bases sous-venantes d'une propriété évaluative survenante comme le danger – qu'elles pourront justifier mon émotion. De ce point de vue, les valeurs en tant qu'objets formels permettent d'apprécier la correction d'une émotion, mais ne sont pas des raisons à proprement parler pour nos émotions.

La théorie réactive, au contraire, affirme que les raisons d'éprouver des émotions sont évaluatives : c'est parce que le chien représente un danger que j'éprouve de la peur en réponse à cette situation. En ce sens, les objets formels des émotions ne constitueraient pas seulement les conditions de corrections de celles-ci, mais seraient aussi nécessaires à leur intelligibilité, un point qui fait d'ailleurs consensus depuis Kenny (1963). Or, si l'on prend ce point au sérieux, soutiennent les partisans de la théorie réactive, alors il faut reconnaître que les objets formels sont aussi des « raisons motivantes » pour les émotions. La peur de S ne fait sens que si S perçoit l'objet de sa peur comme dangereux, et c'est la perception de ce danger qui le pousserait donc à éprouver cette émotion. Comme le dit encore Müller : « L'attribution d'une émotion implique conceptuellement qu'elle est une réponse à son contenu sous la guise de son objet formel » (2017, 288).

Les émotions seraient donc évaluatives non pas en tant qu'elles sont des évaluations (comme le revendique la théorie attitudinale), ni en tant qu'elles sont des perceptions de valeurs (comme le veut la théorie perceptuelle), mais en tant qu'elles consistent en des attitudes de réponse à l'égard d'une conscience de valeur pré-émotionnelle. Il n'y a pas d'appréhension émotionnelle de la valeur à proprement parler, mais seulement une « reconnaissance » (*acknowledgement*) émotionnelle de cette valeur, reconnaissance qui se caractérise par une attitude spécifique. La théorie réactive est donc d'accord avec la théorie perceptuelle sur le fait que l'émotion représente son objet comme étant doté d'une valeur, mais elle ajoute que la présentation de la valeur ne vient pas de l'émotion elle-même. L'émotion est plutôt une reconnaissance de cette valeur déjà appréhendée de façon non-émotionnelle.

Pourquoi admettre que les raisons d'éprouver des émotions seraient évaluatives? N'est-il pas au contraire plus économique d'affirmer que les émotions sont rendues intelligibles par les propriétés non-évaluatives de leurs objets? Massin soulève deux difficultés à cet égard qui semblent plaider en faveur de la théorie réactive contre la théorie attitudinale :

1. Si les émotions étaient justifiées par les propriétés non-évaluatives de leurs objets, alors la plupart des émotions seraient généralement injustifiées étant donné que chaque objet possède de nombreuses propriétés non-évaluatives susceptibles de justifier nos émotions, et dont nous ne sommes pas conscients : ce sont non seulement les crocs du chien qui justifient ma peur, mais son comportement, son poids, ou encore le fait que je n'aie aucun moyen de me défendre, qu'il ne soit pas attaché, etc. Toutes ces propriétés constituent en effet les bases sous-jacentes sur lesquelles la valeur du danger est censée survenir. Mais il semble improbable que nous puissions être conscients de l'ensemble de ces propriétés non-évaluatives à chaque fois que nous éprouvons une émotion. Ce serait une implication surprenante de la théorie attitudinale, que la théorie réactive de son côté évite facilement en affirmant que les raisons d'éprouver des émotions sont en réalité évaluatives. Notons cependant que cette stratégie revêt un coût relativement élevé, puisqu'elle présuppose qu'il y a bien des états de choses évaluatifs, qui sont par ailleurs appréhendés par une faculté pré-émotionnelle dont la nature reste à préciser.
2. Dire que les émotions sont justifiées par des propriétés naturelles ou non-évaluatives revient de toute façon à faire appel implicitement à des justifications évaluatives. C'est parce que les crocs du chien sont dangereux qu'ils justifient ma peur. Au mieux, il y a donc une relation d'implication tacite entre les propriétés naturelles et les propriétés évaluatives. Et là où la théorie attitudinale affirme que nous allons tenter de justifier nos émotions en nous appuyant sur les propriétés naturelles de leurs objets, Massin affirme au contraire que nous commençons par justifier nos émotions sur la base d'une attribution de valeur, et que nous cherchons ensuite, en examinant les propriétés naturelles de leurs objets, à justifier cette attribution de valeur préalable.

Comme le souligne Massin, ces arguments ne démontrent pas la vérité de la théorie réactive, mais ils suggèrent que cette théorie serait en meilleure posture que la théorie attitudinale en regard de la question de la justification des émotions.

3.2 DES RAISONS DE DÉFENDRE L'ATTITUDINALISME

La théorie attitudinale est-elle en mesure de répondre à ces deux difficultés? Commençons par noter qu'elle partage avec la théorie réactive une même conception de l'émotion, en tant que réponse à un objet ou une situation plutôt que comme un genre de réceptivité, comme le voudrait par exemple la

théorie perceptuelle ainsi que d'autres théories cognitives classiques. Mais de ce point de vue, comme l'ont souligné récemment Deonna et Teroni (2024a), la théorie réactive semble appauvrir la nature des émotions, en les présentant systématiquement comme un type de reconnaissance des valeurs, là où l'on pourrait être tenté d'identifier pour chaque « famille » d'émotions des manières spécifiques de s'engager avec des familles de valeurs : les émotions de la famille de la colère consisteraient ainsi à s'engager de façon hostile avec les valeurs correspondantes (l'offense, l'obstruction, etc.), et les émotions de la famille de la peur seraient des manières de s'engager avec les valeurs qui leur sont liées (le danger, le terrifiant, etc.)⁷.

L'idée serait donc qu'il y ait pour chaque famille d'émotion une famille de valeurs correspondantes sur le plan intentionnel, et sur le plan phénoménologique des familles de réactions spécifiques consistant en une façon de s'engager avec ces valeurs, que ce soit par l'hostilité, par l'évitement, ou encore par l'inspection, le réconfort, et bien d'autres types d'attitudes possibles. Ce qui distingue alors ces familles d'attitude les unes des autres, c'est la façon dont elles nous poussent à nous engager avec leur objet, de sorte que même si certaines émotions partagent parfois des tendances à l'action communes (on peut vouloir détruire l'objet de notre émotion aussi bien dans la colère que dans la peur par exemple), leur expérience en tant que manière de s'engager avec un objet n'est pas organisée de la même façon (la colère est organisée autour de l'hostilité alors que la peur est organisée autour de l'évitement).

Faut-il admettre pour autant, comme les partisans de la théorie réactive, que les émotions supposent une conscience pré-émotionnelle de la valeur? Cette concession risque de réintroduire une forme de représentation de valeur dans les émotions, ou plus précisément encore, préalablement aux émotions. Or, l'une des revendications centrales de la théorie attitudinale est que les émotions ne sont pas des représentations de valeurs. Et même en admettant qu'il puisse y avoir une représentation de valeur préalable à l'émotion, cela pose le problème de savoir d'où viendrait cette représentation, et sous quelle forme elle pourrait avoir lieu. D'une certaine façon, ces deux questions sont liées, puisqu'en expliquant sous quelle forme l'appréhension pré-émotionnelle de valeur a lieu, on pourra désigner en même temps la faculté dont elle procède. En admettant par hypothèse qu'une telle appréhension ait lieu, Deonna et Teroni soulignent qu'elle ne peut vraisemblablement pas être de nature conceptuelle, au sens où l'on reconnaîtrait que l'objet de notre émotion tombe sous un certain concept de valeur, car on serait alors bien en peine d'expliquer la vaste palette d'émotions que certaines créatures peuvent éprouver indépendamment de toute maîtrise conceptuelle. La solution que proposent les réactivistes de leur côté, consiste à dire que les valeurs sont identifiées

7. L'approche des émotions en termes de famille est relativement courante. Je la présente dans LNE, p. 85-86.

par leur profil phénoménologique, c'est-à-dire par un certain ressenti propre à chaque type de valeur. Il faudrait donc postuler ici des sentiments de valeur spécifiques, ou un type d'intuition susceptible de nous donner accès à ces valeurs⁸.

En regard de ces deux approches – conceptuelle et intuitionniste – Deonna et Teroni proposent ce qu'ils nomment le « modèle simple », selon lequel nous n'avons pas réellement de conscience pré-émotionnelle des valeurs, mais plutôt une conscience pré-émotionnelle des porteurs de valeurs, c'est-à-dire des propriétés naturelles qui constituent typiquement une instance de valeur donnée. Il n'y aurait donc pas d'accès conceptuel ni phénoménal à la valeur, ici, puisqu'il n'y a pas d'accès direct à celle-ci, mais seulement un accès indirect via leurs bases sous-venantes. En suivant cette voie, l'attitudinaliste peut alors tenter de répondre aux différents défis soulevés par le réactiviste, et notamment aux arguments mobilisés par Massin.

Comment pouvons-nous identifier les propriétés non-évaluatives pertinentes pour nos émotions, étant donné le nombre indéfini de propriétés susceptibles d'être ici pertinentes? La réponse de Deonna et Teroni est que nous sommes généralement capables d'identifier non pas des propriétés naturelles isolément les unes des autres, mais des constellations de propriétés naturelles typiquement susceptibles d'être les bases sous-venantes de certaines valeurs. Ce point peut d'ailleurs être rendu intelligible aisément il me semble : il est clair que nous avons une certaine familiarité avec la façon dont certains objets et situations sont susceptibles d'instancier typiquement certaines valeurs, et qui entrent dans ce que Prinz nomme les « fichiers de calibration » de nos émotions. Ces fichiers contiennent en effet l'ensemble des *stimuli* susceptibles de déclencher typiquement nos émotions : c'est ainsi que le comportement d'un chien aussi bien que l'éventualité d'une guerre ou d'une récession économique peuvent être des déclencheurs typiques de la peur, ou que les victoires, les prouesses et la possession de certains objets rares peuvent être des déclencheurs typiques de la fierté.

Il y a ainsi un genre de familiarité avec les déclencheurs typiques de nos émotions qui facilite leur identification comme potentiels porteurs de certaines valeurs, qui correspond à ce que l'on appelle typiquement notre « sensibilité » (Deonna et Teroni, 2022). Cette familiarité avec des déclencheurs ou des constellations de déclencheurs typiques, n'a rien de mystérieux, et c'est à elle que fait appel Kahneman (2011) lorsqu'il cherche à expliquer la distinction entre le fonctionnement du système 1 et celui du système 2 en psychologie. En effet, Kahneman prend l'exemple d'un pompier qui, en plein milieu d'une intervention, déclare qu'il faut sortir du bâtiment sans savoir exactement pourquoi, juste avant que le bâtiment ne s'effondre, ou d'un joueur d'échec qui doit prendre des décisions rapides en identifiant des opportunités d'action. Dans ce genre de situations, nous nous basons en effet sur des

8. Voir notamment Boisserie-Lacroix (2022), et Naar (2023).

constellations de propriétés que nous identifions comme susceptibles d'instancier certaines valeurs, quoique nous ne soyons pas nécessairement en mesure de les verbaliser spontanément.

Inversement, il arrive qu'il nous soit difficile d'identifier des constellations typiques, ce qui peut engendrer un genre de flou émotionnel, tant que nous ne sommes pas parvenus à évaluer la situation. Quoiqu'il en soit, l'analyse esquissée ici permet aussi de répondre au défi proposé par Delamare d'expliquer les attitudes évaluatives froides : je peux en effet très bien réagir sans émotion à ce que j'identifie comme étant des constituants d'une valeur donnée, du fait de ma sensibilité. Le fait de réagir froidement ou chaudement face à une même constellation de propriétés ne dépend alors que de nos bases motivationnelles : je peux être indifférent à la présence de certaines valeurs, ou inversement avoir un attachement à celles-ci qui me disposera à réagir émotionnellement.

Ici le partisan de l'approche réactive pourrait être tenté de souligner que Kahneman s'appuie précisément sur les exemples précédents pour montrer que nous avons accès à certaines informations de façon « intuitive », de sorte qu'en s'appuyant sur ces exemples le partisan de l'attitudinalisme pourrait bien donner involontairement du grain à moudre à l'idée que les émotions reposeraient sur une intuition de valeurs. Mais le nerf du débat porte moins sur l'appel ou non à l'intuition (qui n'est au fond que le nom que donne Kahneman au fonctionnement rapide de la pensée, par opposition à son fonctionnement lent et réfléchi), que sur l'idée d'une saisie pré-émotionnelle des valeurs. Or ici il n'y a pas de controverse : l'attitudinaliste soutient que nous n'avons en aucun cas un accès aux valeurs, mais seulement à leurs porteurs, et l'analyse de Kahneman ne permet clairement pas de trancher ce point.

La deuxième difficulté soulevée par Massin est moins embarrassante, car c'est précisément une revendication de la théorie attitudinale qu'il puisse y avoir une relation d'implication entre les propriétés naturelles et les propriétés évaluatives, de sorte que les secondes ne nous sont données qu'indirectement par l'entremise des premières. Massin, conteste ce point, et soutient que l'on va des propriétés évaluatives aux propriétés non évaluatives lorsque nous cherchons à justifier nos attributions pré-émotionnelles de valeurs. Mais dans ce cas, les émotions seront désormais justifiées par un sentiment évaluatif, qui devra à son tour être justifié par des propriétés non évaluatives. Il semble que l'on ait simplement ajouté une étape intermédiaire superflue dans la justification des émotions.

3.3 OBJETS FORMELS ET REPRÉSENTATION DE VALEURS

Le point précédent pose la question de savoir quel rôle jouent les objets formels du point de vue de la théorie attitudinale. En effet, selon celle-ci, l'objet formel d'une émotion est la valeur qui permet d'apprécier la correction de cette émotion. Mais cet objet formel ne figure pas, selon la théorie attitudinale, dans le contenu de l'émotion, sans quoi l'émotion aurait un contenu

évaluatif et ne serait rien d'autre qu'une représentation de valeur – comme le soutiennent la théorie perceptuelle et d'autres théories appartenant à ce que Delamare nomme le modèle de la « cognition axiologique ». En affirmant néanmoins que les émotions peuvent être justifiées par leurs bases cognitives, c'est-à-dire par les propriétés naturelles de leurs objets qui constituent les bases sous-venantes d'une valeur donnée, l'approche attitudinale semble néanmoins impliquer que les émotions pourraient être des représentations de valeurs. Auquel cas l'approche attitudinale se réduirait en réalité à une approche représentationnelle, de sorte qu'en renonçant à la théorie réactive, elle basculerait nécessairement vers une forme de représentations des valeurs (voir notamment Müller, 2017, pour une critique de ce type).

De plus, comment comprendre la correction d'une émotion, si ce n'est en termes de correspondance de son contenu avec la valeur qui constitue son objet formel ? Rossi et Tappolet (2019) soutiennent à cet égard que l'attitudinalisme est condamné à admettre que le contenu des émotions est axiologique en réalité, et que les émotions ne sont rien d'autre que des représentations de valeurs. En effet, selon eux, l'objet formel d'un état mental fait en principe partie de son contenu. Si tel est le cas, alors la conception perceptualiste des objets formels serait préférable à la conception attitudinaliste. C'est un point que contestent Deonna et Teroni (2022) en soulignant que l'objet formel ne fait pas toujours partie du contenu d'un état mental : la vérité par exemple ne fait pas partie du contenu d'une croyance, même si les croyances visent la vérité. Mais il faut reconnaître avec Rossi et Tappolet que c'est peut-être le cas pour d'autres types d'états mentaux : la couleur, par exemple, entre dans le contenu des perceptions chromatiques tout en étant leur objet formel. Dans ce cas, la question est de savoir si un objet formel peut nous permettre d'évaluer la correction d'une émotion en termes de correspondance de son contenu avec la valeur que vise cette émotion, sans pour autant figurer dans le contenu de cette émotion.

Plusieurs réponses sont possibles ici. La première est que l'on peut tracer une distinction entre les états mentaux réceptifs comme la perception, et les états mentaux « réactifs » comme le sont les émotions et les croyances, qui manifestent une forme de sensibilité aux preuves en leur faveur (Deonna et Teroni, 2022). En ce sens, la correction de ces états réactifs semble être intrinsèquement liée aux types d'attitudes qui les caractérisent, et la normativité des émotions semble dériver précisément du fait que ces dernières consistent à traiter certaines propriétés comme méritant certaines réponses comportementales. De ce point de vue, il y aurait au moins une raison de penser que les émotions ne peuvent pas être assimilées à des états perceptifs, et que leurs conditions de correction ne fonctionnent pas sur le modèle de ces derniers. La normativité des émotions, tout comme celle des croyances, proviendrait précisément du fait que ce sont des attitudes sensibles à certaines preuves (liées aux valeurs pour les premières, et à la vérité pour les secondes).

Une deuxième réponse, complémentaire, est que si l'objet formel devait faire partie du contenu des émotions, l'analyse des concepts de valeurs à

partir de nos concepts d'émotions risquerait d'être circulaire (Deonna et Teroni, 2021). Le concept de danger, par exemple, renverrait au concept d'une émotion de peur appropriée, qui ne serait elle-même rien d'autre que la représentation du danger, de sorte que le concept de peur renverrait au concept d'une expérience du danger. Du point de vue de la théorie attitudinale, en revanche, le concept d'une émotion appropriée peut se comprendre de façon non circulaire en termes de réponses comportementales appropriées en regard de certaines constellations de propriétés naturelles, sans présupposer des concepts de valeurs.

Il y a donc de bonnes raisons, me semble-t-il, de continuer à préférer la théorie attitudinale aux théories réactives et aux théories perceptualistes ou représentationalistes. En particulier, il me semble qu'un point commun à ces deux dernières théories est qu'elles présupposent que le lien entre émotions et valeurs serait déjà donné dans le concept d'émotion, alors qu'il semble plus vraisemblable que les émotions nous permettent d'éclairer les concepts de valeurs à partir des types d'attitudes qui les caractérisent. Comme je le soulignais déjà dans LNE, une idée maîtresse de l'approche attitudinale est que les émotions ne nous révèlent pas des valeurs, mais que c'est en cherchant à les justifier que l'on peut tenter d'élucider les concepts de valeurs. En renonçant à l'idée que les valeurs font partie du contenu des émotions, on perd certes un aspect intéressant des approches rivales, mais on aboutit à une image qui me semble empiriquement plus fidèle à nos pratiques évaluatives.

3.4 CONCLUSION

Si je reconnais l'importance du correctif majeur que propose Delamare, il me paraît préférable de conserver la version de la théorie attitudinale telle quelle sans effectuer sa mise-à-jour, comme il le propose, vers la théorie réactive. À mon sens, la théorie attitudinale reste la théorie des émotions la plus modeste en termes d'engagements ontologiques et conceptuels : elle ne présuppose pas l'existence de valeurs, ni même des concepts de valeurs dans l'analyse des émotions appropriées. Je serais donc plutôt heureux de laisser cet aspect de LNE en l'état.

4 PROBLÈMES DE JUSTIFICATION ET ÉMOTIONS INCONSCIENTES. RÉPONSE À ARTÛRS LOGINS

J'ai dit dans l'introduction de ces réponses que les discussions concernaient souvent les points qui me paraissaient les plus fragiles de LNE. Le commentaire d'Artûrs Logins ne fait pas exception à cette règle, et cible un aspect sur lequel je m'attendais à ce que les épistémologues soient particulièrement sourcilleux (celui qui porte sur la nature de la justification telle que la comprend l'approche motivationnelle), mais aussi un point sur lequel je m'y attendais beaucoup moins (la question des émotions inconscientes), même si ce dernier ne m'a jamais complètement satisfait. J'essaierai d'abord

de défendre, autant que je puisse le faire, la conception de la justification élaborée dans LNE, avant de m'attarder plus longuement sur les émotions inconscientes.

4.1 LA JUSTIFICATION DES ÉMOTIONS EN QUESTION

Dans la première partie de son commentaire, Artūrs Logins jette un doute sur un aspect de la stratégie employée dans LNE pour défendre la théorie motivationnelle, relatif à la justification des émotions. En effet, une implication potentiellement problématique de la théorie motivationnelle est que les motivations sont susceptibles de pouvoir contribuer à la justification de nos émotions, puisqu'elles constituent les valeurs auxquelles celles-ci répondent. Pourtant, de façon assez intuitive, on pourrait être tenté de dire que les motivations expliquent nos émotions mais qu'elles ne les justifient pas. Mon désir de passer du temps avec mes amis explique la joie que j'éprouve en leur présence. Mais celle-ci semble se justifier par la qualité du moment que nous passons ensemble, et non pas par mon désir. Comme le souligne Logins, admettre que nos motivations puissent justifier nos émotions s'accompagne d'un risque d'arbitrarité : on pourrait être tenté de soutenir en effet qu'il suffit d'avoir une certaine motivation pour justifier une émotion donnée, quand bien même celle-ci ne serait apparemment pas justifiée par son objet. Cela semble improbable : le fait que j'aie le désir de passer un moment avec mes amis ne suffirait pas à justifier ma joie, si nous sommes en réalité en train de passer une journée catastrophique.

Une première stratégie que j'ai tenté d'employer dans LNE consistait à soutenir que dans une approche contextualiste, les mêmes raisons explicatives pouvaient devenir des raisons justificatives, mais comme le souligne Logins, c'est une stratégie insuffisante, car cela ne permet pas de soutenir qu'une même raison peut être explicative et justificative. Ce qu'il reste à expliquer dans tous les cas, ce sont les raisons pour lesquelles les motivations pourraient compter comme des raisons justificatives. Logins pense qu'à partir d'une définition claire des raisons justificatives, on peut douter que les motivations puissent jouer ce rôle. Les raisons justificatives, en effet, sont censées désigner des considérations qui doivent pouvoir jouer le rôle d'inférences dans le raisonnement d'un sujet en faveur de son attitude. C'est ce qu'on pourrait appeler la contrainte normative sur la justification : il faut que ces considérations soient logiquement pertinentes en regard du caractère approprié de l'attitude. Par ailleurs, ces considérations doivent également être accessibles à ce même sujet. C'est ce qu'on pourrait appeler la contrainte de plausibilité psychologique sur la justification : il faut que le sujet puisse être conscient de ces considérations, sans quoi il ne saurait pouvoir les mobiliser.

À partir de là Logins considère qu'il est peu plausible que les motivations puissent jouer un rôle justificatif. Il me semble pourtant que certaines motivations (dont le désir, les sentiments, ou encore les préférences) sont précisément susceptibles de satisfaire ces deux contraintes : nous sommes gé-

néralement conscients de celles-ci, et il semble pertinent de les mentionner comme bases pour nos émotions. Mon désir de réussir une voie difficile en escalade n'explique pas seulement ma joie, il contribue à la justifier au sens où je n'aurai probablement pas de raison de me réjouir sans lui. Ce n'est pas simplement une raison explicative qui rendrait intelligible ma joie, mais c'est une raison normative au sens où elle légitime au moins partiellement ma joie. Si on me demande pourquoi je suis content, une partie de la réponse est que je suis satisfait de mon accomplissement, et une autre partie est que j'y tenais. Et si je ne me réjouissais pas, mes proches pourraient me faire remarquer, étonnés, que je devrais me réjouir !

Cette contribution des états motivationnels à la justification, selon LNE, est du moins légitime si ma motivation est elle-même justifiée, ce qui permet par ailleurs d'échapper au problème de l'arbitrarité mentionné plus haut. Notons que ce modèle semble d'autant plus acceptable si l'on admet, comme le suggère Rioux (voir la section 2), une contrainte de proportionnalité dans la justification des émotions. Le fait que j'aie un désir intense de réussir cette voie constitue une bonne raison d'éprouver une joie intense, au-delà de la difficulté intrinsèque de la voie.

Une difficulté possible, sur laquelle LNE passait trop rapidement, tient peut-être au fait que nos motivations, malgré tout, ne suffisent pas à justifier nos émotions, mais peuvent seulement contribuer à cette justification. Pour reprendre l'exemple précédent, mon désir de réussir contribue à justifier ma joie, mais ce qui la justifie en dernier lieu, c'est ma réussite et le fait d'avoir ainsi satisfait mon désir. Dans ce cas, on pourrait se demander à quoi peut bien servir la justification par les motivations, si elle n'est pas suffisante ? Est-elle seulement nécessaire ? Et inversement, la justification de l'émotion par des faits objectifs n'est-elle pas suffisante ? C'est ce que semble suggérer Logins en conclusion de son premier point. Dans LNE, j'introduisais en réponse à ce genre de préoccupation l'idée d'une « stratification de la justification », inspirée par une forme de faillibilisme, en proposant une distinction entre justification primaire et justification secondaire, qui me semble malgré tout insuffisante, et sur laquelle j'aimerais revenir ici.

L'idée de cette distinction était que les motivations font l'objet d'une justification primaire, au sens où cette justification est première, et ne repose pas sur la justification d'autres états affectifs, mais seulement sur les bonnes raisons que nous avons d'avoir une motivation donnée. Les émotions, de leur côté, font l'objet d'une justification secondaire, au sens où elles sont justifiées à la fois par les motivations, et par les propriétés de leurs objets. Mais une difficulté importante tient au fait que je suggère régulièrement que les motivations semblent souvent suffire à justifier nos émotions, ce qui paraît simplement faux. Comme l'a remarqué Bonard (2024), cette thèse laisserait ouverte la possibilité à une forme de cohérentisme que je souhaitais pourtant rejeter, où l'émotion raciste d'un nazi cohérent serait justifiée parce qu'elle est cohérente avec au moins l'une de ses motivations. Je ne prends clairement pas suffisamment au sérieux cette objection dans LNE, en soutenant

que les racistes savent très bien que leur idéologie repose sur des croyances infondées. Après tout, le contraire est tout à fait plausible également. Et si tel est le cas, la thèse motivationnelle est en danger car on pourrait être tenté de penser qu'un nazi cohérent pourrait avoir effectivement de bonnes raisons, de son point de vue, d'éprouver les émotions qui sont les siennes.

En réponse à ces difficultés, il me semble que l'on peut tenter de modifier la théorie motivationnelle de plusieurs façons, que je me contenterai d'esquisser ici, et qui sont par ailleurs potentiellement complémentaires :

- 1) Admettre que les motivations sont nécessaires mais non suffisantes à justifier nos émotions, tout comme les propriétés de l'objet sont nécessaires et non suffisantes. Ceci conduit à une thèse étrange à première vue, puisque cela revient à dire que les émotions sont justifiées par deux types de sources différentes et complémentaires. Mais après tout, ce n'est pas si absurde : réussir à un concours dont je ne me soucie pas ne me donne qu'une demi-raison de me réjouir, faute d'avoir la motivation appropriée. Et l'inverse semble vrai également : me soucier de ma santé me donne une demi-raison de m'inquiéter d'être malade, pour autant que j'aie par ailleurs des raisons objectives de croire que je pourrai tomber malade. Ainsi, les sources de justification viennent à la fois de nos motivations et des faits extérieurs et l'émotion ne peut être considérée comme correcte que si ces deux sources parlent en sa faveur.
- 2) Renoncer clairement à l'internalisme de la justification qui affleure régulièrement dans certaines formules de LNE, lorsque j'aborde le cas du nazi cohérent, bien que j'essaie par ailleurs plus explicitement de défendre une forme de fiabilisme. Une variante externaliste de la théorie motivationnelle pourrait ainsi stipuler que les émotions tout comme les motivations sont justifiées lorsqu'elles reposent sur des sources de justifications qui sont elles-mêmes fiables ou plus radicalement sur des faits qui parlent en leur faveur.
- 3) Dans le cas des émotions morales, il me semble toutefois que les choses sont un peu plus compliquées, car les motivations qu'il conviendrait d'avoir dépendent de considérations qui sont à la fois liées à certaines de nos motivations (comme notre désir de minimiser les souffrances inutiles, ou de maximiser les expériences plaisantes), ainsi que de règles qui font l'objet d'une construction. Le nazi cohérent à cet égard, passe simplement à côté des considérations qui devraient l'amener à réviser ses motivations incorrectes, même s'il a des bonnes raisons, de son point de vue, d'avoir de telles motivations. Est-ce que cela revient à accepter une forme de réalisme au sujet des valeurs ? C'est possible, mais comme je l'ai rappelé dans la partie précédente consacrée aux objections de Catherine Rioux, la théorie motivationnelle est de toute façon compatible avec plusieurs variétés de réalisme.

Ces suggestions impliquent probablement des modifications substantielles de certains aspects de la théorie motivationnelle, mais explorer celles-ci est

clairement un travail qui sort du cadre de cette réponse, et que j'espère avoir d'autres occasions de mener.

4.2 N'Y A-T-IL VRAIMENT PAS D'ÉMOTIONS INCONSCIENTES ?

Le deuxième point de Logins concerne mon rejet des émotions inconscientes. Bien que la section consacrée à ce point dans LNE soit relativement courte, elle masque un certain nombre de doutes que j'ai toujours entretenus sur cette question. Je saute donc avec plaisir sur l'occasion que me présente Artūrs Logins de reprendre cette question.

Si l'on suit la critique de Logins, on est apparemment contraint d'admettre qu'il y a des émotions inconscientes dès lors que l'on accepte l'argument anti-luminosité de Williamson. Selon cet argument, en effet, la connaissance n'implique pas de dimension réflexive : savoir quelque chose au sujet de x n'implique pas en plus de savoir que nous avons ce savoir au sujet de x . Appliqué aux émotions, ce principe devrait nous conduire à admettre que lorsque nous sommes en train d'éprouver une émotion, nous ne savons pas nécessairement que nous l'éprouvons. Cela revient donc à dire qu'on peut éprouver une émotion sans en être conscient.

Comment interpréter cette conclusion ? Logins suggère deux pistes. La moins convaincante, selon lui, serait de dire que l'on peut être conscient de l'émotion d'un point de vue phénoménal, au sens où il y aurait un *quale* caractéristique pour cette émotion, faisant l'objet d'un ressenti, mais que l'on pourrait en même temps ne pas savoir que l'on ressent cette émotion. Mais cette distinction entre conscience phénoménale et savoir est peu plausible selon lui, pour des raisons qui me semblent également convaincantes. Une deuxième manière d'interpréter cette conclusion est qu'elle devrait nous inciter à penser que la phénoménologie des émotions n'est pas si centrale, et qu'il peut y avoir des émotions qui ne sont pas ressenties.

Cette conclusion est-elle vraisemblable ? Avant de poursuivre la discussion, notons que d'autres philosophes ont récemment soutenu que les émotions pourraient effectivement être inconscientes dans certains cas. Lemaire (2022) affirme par exemple que les émotions peuvent soit avoir un objet intentionnel inconscient⁹, soit ne pas avoir de phénoménologie identifiable, en s'appuyant à la fois sur des exemples ordinaires et sur des travaux empiriques. D'un point de vue intentionnel, il peut arriver que nous ayons du mal à identifier l'objet de nos émotions (un point sur lequel LNE est en accord pour rappel), et des expériences montrent que des patients dont l'attention visuelle est lésée peuvent répondre émotionnellement à des objets sans être conscients de ces objets. D'un point de vue phénoménal, des recherches sug-

9. Bonard (2022), dans une veine similaire, affirme que certaines humeurs pourraient être inconscientes au sens où l'objet intentionnel des humeurs ne nous serait pas accessible, de sorte que l'intentionnalité de ces humeurs ne ferait pas partie de leur caractère phénoménal.

gèrent que des *stimuli* subliminaux peuvent engendrer des comportements qui sont émotionnellement influencés, bien qu'aucune émotion ne semble ressentie, et l'on peut aussi imaginer que certaines émotions puissent être suffisamment faibles pour ne pas être ressenties dans la vie courante. C'est donc une approche qui pourrait aller dans le sens de la conclusion de Logins.

Dans un article qui s'efforce de faire la synthèse des travaux récents sur la possibilité qu'il y ait des émotions inconscientes, Sarah Arnaud (2023) propose de son côté de distinguer, sous la forme d'un trilemme, trois propositions qui ont chacune une certaine plausibilité, mais que l'on ne peut faire tenir simultanément :

1. « Les émotions peuvent être inconscientes »
2. « Les émotions ont par essence une phénoménologie »
3. « La phénoménologie émotionnelle implique la conscience émotionnelle »

La proposition 1 comme on vient de le voir est rendue plausible par différents travaux empiriques (Arnaud en ajoute bien d'autres du même type), et il est par ailleurs concevable qu'on puisse être joyeux ou en colère sans s'en rendre compte. La proposition 2 semble également inévitable dans la mesure où une émotion sans phénoménologie ressemblerait à une simple évaluation froide ou, comme le dit Arnaud, à un jeu d'acteur. Dans ce cas, la question se pose de savoir où l'on devrait classer les cas rapportés par Lemaire, mais il me semble qu'ils renvoient davantage à des situations où la phénoménologie peut être présente sans être ressentie – parce qu'elle est trop faiblement prononcée par exemple – si bien qu'ils concerneraient davantage la proposition 3. Or la proposition 3 ressemble également à un truisme : par principe la phénoménologie semble être consciente, puisque si l'on ressent quelque chose, c'est que l'on en fait l'expérience, et que l'on doit donc d'une certaine façon être en mesure d'avoir un accès conscient à cette chose.

À première vue, il n'est pas évident de savoir si Logins propose de rejeter 2 ou 3, mais j'ai tendance à penser que sa suggestion concerne davantage 3. De mon côté, je ne rejette pas complètement 1 dans LNE, mais j'affirme plus modestement que certains cas d'émotions apparemment inconscientes sont des cas dans lesquels nous ne prêtons pas attention à notre ressenti affectif, ou dans lesquels nous ne parvenons pas à le catégoriser adéquatement, ce qui renvoie davantage à la proposition 3 également. Ces deux types de cas, quoi qu'il en soit, sont compatibles avec l'idée que l'on pourrait avoir du mal à identifier l'objet intentionnel de nos émotions, ou à percevoir sa phénoménologie.

Comme le souligne Arnaud, rejeter complètement 1 reviendrait à nier la possibilité que les enfants ou certains animaux puissent éprouver des émotions. Si l'on rejette 3, on risque de se priver des moyens de distinguer une émotion en tant que telle de simples réactions corporelles automatiques ou réflexes qui ressemblent à cette émotion, que l'on peut retrouver dans les rires ou les pleurs pathologiques. Arnaud propose alors de rejeter une certaine interprétation de 3 : si la phénoménologie de l'émotion désigne bien

un état conscient, cet état n'implique pas de réflexivité. Autrement formulée, la solution que propose Arnaud revient à dire que l'on peut très bien être conscient au sens où notre attention est portée sur un objet, sans se rendre compte que l'on est dans cet état émotionnel. Arnaud soutient que la conscience émotionnelle dans ce cas est « une expérience de l'objet formel de l'émotion », qui pourrait ne pas être représenté consciemment. Selon elle, on serait donc émotionnellement conscient de l'objet de notre émotion (ou de certains sons, ou de certaines sensations), sans être conscient de notre expérience émotionnelle comme telle, ni du fait que l'on attribue un objet formel à l'objet de notre émotion.

Pour des raisons exposées précédemment dans ma réponse à Alexis Delamare, ainsi que dans le chapitre IV de LNE, il me paraît cependant douteux que l'expérience émotionnelle soit généralement celle d'un objet formel. Il me semble plus probable que l'expérience émotionnelle, dans les cas d'émotions inconscientes, soit avant tout une expérience valencée positivement ou négativement au sujet d'un objet, qui peut se traduire par une certaine orientation de nos pensées ou par notre attitude corporelle. En revanche, il me semble clair qu'une telle expérience n'implique pas nécessairement de conscience réflexive sur elle-même.

À la lueur de ces analyses, on peut faire au moins deux remarques. D'une part, il me semble que la thèse que défend Arnaud permet à la fois de faire sens de l'argument de Logins (oui, il y a des émotions inconscientes) tout en résistant à l'idée que les émotions n'ont pas nécessairement de phénoménologie. Il me semble que toute affirmation à ce sujet revient seulement à dire que la phénoménologie n'a pas nécessairement de dimension réflexive, et cela semble être dans la droite ligne de l'argument anti-luminosité. Logins rétorquerait peut-être (comme le suggèrent les dernières lignes qui précèdent sa conclusion), que si l'on n'est pas conscient d'avoir une émotion, alors on n'a pas d'expérience consciente non plus des propriétés de l'objet, auquel cas c'est bien la thèse 2 qu'il rejette avec 3. Mais c'est un point qui me paraît plus délicat à défendre, car comme nous l'avons vu, les exemples emblématiques d'émotions considérées comme inconscientes impliquent en principe une orientation de notre attention sur certaines propriétés. Même lorsque cela concerne des *stimuli* subliminaux, il y a bien une expérience consciente malgré tout. Il me semble qu'il faudrait un argument supplémentaire pour nier ce point.

D'autre part, il me semble que ces analyses s'inscrivent au fond dans la lignée des distinctions que je proposais dans LNE. Il est clair que ces distinctions restent assez sommaires, et qu'elles donnent peut-être l'impression que je nie simplement qu'il y ait des émotions inconscientes (certaines formules me paraissent à cet égard un peu maladroites rétrospectivement). Mais ce n'est pas exactement le cas. Je soutiens en effet que l'on peut avoir un ressenti émotionnel, et même une évaluation émotionnelle, sans que l'on ne s'en rende compte, et ceci me paraît compatible avec les cas d'émotions inconscientes que j'ai mentionnés ici.

4.3 CONCLUSION

La partie épistémologique de LNE liée à la justification des émotions m'a toujours paru la plus fragile, et Artūrs Logins n'a pas manqué d'en pointer les difficultés les plus manifestes. J'espère que certaines des solutions que je propose ici pour corriger la thèse motivationnelle pourront donner lieu à des versions plus heureuses de cette approche. Quant à la question des émotions inconscientes, je crois que la discussion mériterait d'être poursuivie pour que l'on puisse identifier plus précisément un potentiel lieu de désaccord. Mais le point central me paraît ici être lié à la nature de la phénoménologie, et aux propriétés que l'on estime être nécessaires pour celle-ci. Je serai heureux que la discussion puisse également progresser sur ce point et, si c'est le cas, apporter un nouveau correctif aux thèses défendues dans LNE.

BIBLIOGRAPHIE

- Arnaud, Sarah (2025). « Unconscious Emotions ». *Erkenntnis* 90 (1) :285-304.
- Bonard, Constant (2022). « The rationality of mood ». J. Deonna, C. Tappolet and F. Teroni (Eds.), *A Tribute to Ronald de Sousa*. URL : <https://sites.cisa-unige.ch/ronald-de-sousa/>.
- (2024). « La nature des émotions : Une recension partisane ». *Les ateliers de l'éthique / The Ethics Forum* 18 (1) :29-49.
- Brady, Michael (2013). *Emotional insight : the epistemic role of emotional experience*. New York : Oxford University Press.
- Boisserie-Lacroix, Céline (2022). « Une perspective extrinsèque sur les normes émotionnelles ». *Revue de métaphysique et de morale*, 114(2), 173-188.
- Callard, Agnes (2018). *Aspiration: The Agency of Becoming*. New York : OUP.
- Deonna, Julien A. et Teroni, Fabrice (2012). *The Emotions: a Philosophical Introduction*. New York : Routledge.
- (2021). « Which Attitudes for the Fitting Attitude Analysis of Value? » *Theoria* 87 (5) :1099-1122.
- (2022). « Why are emotions epistemically indispensable? » *Inquiry*, 68(1), 91-113.
- (2024). « Emotions and Their Correctness Conditions: A Defense of Attitudinalism ». *Erkenntnis* 89, 45-64.
- Kahneman, Daniel (2011). *Thinking, Fast and Slow*. New York : New York : Farrar, Straus and Giroux.
- Kenny, Anthony (1963). *Action, Emotion and Will*. NY : Humanities Press.
- Lemaire, Stéphane (2022). « The intentionality of emotions and the possibility of unconscious emotions ». J. Deonna, C. Tappolet and F. Teroni (Eds.), *A Tribute to Ronald de Sousa*. URL : <https://sites.cisa-unige.ch/ronald-de-sousa/>.
- Lepine, Samuel (2016). « Les découvertes émotionnelles et les motivations affectives ». *Revue philosophique de la France et de l'étranger*, Tome 141(2), 179-196.
- (2022). « On salience, confabulation, and emotion's reliability ». J. Deonna, C. Tappolet and F. Teroni (Eds.), *A Tribute to Ronald de Sousa*. URL : <https://sites.cisa-unige.ch/ronald-de-sousa/>.
- Mason, Sam (2025). « Dimensions of Emotional Fit ». *The Philosophical Quarterly* 75 (1) :125-146.
- Massin, Olivier (2023). « The reactive theory of emotions ». *European Journal of Philosophy* 31 (3) :785-802.
- Müller, Jean Moritz (2017). « How (Not) to Think of Emotions as Evaluative Attitudes ». *Dialectica* 71 (2) :281-308.

- Naar, Hichem (2023). « Value Feelings: A Defense », *Philosophies* 8 (4) :69.
- Oatley, K., & Johnson-Laird, P. N. (2002). « Emotion and reasoning to consistency ». In S. C. Moore & M. Oaksford (Eds.), *Emotional cognition* (pp. 157–181). Johns Benjamins.
- Paul, Laurie (2014). *Transformative Experience*. Oxford, GB : Oxford University Press. (Pour la traduction : 2023, *Ces Expériences qui nous transforment*, trad. S. Dunand et S. Réhault, Paris, Éliott, coll. "La part des choses").
- Rossi, Mauro et Tappolet, Christine (2019). « What kind of evaluative states are emotions? The attitudinal theory vs. the perceptual theory of emotions ». *Canadian Journal of Philosophy* 49 (4) :544-563.
- Scarantino, Andrea (2014). « The motivational theory of emotions ». In Justin D'Arms Daniel Jacobson, *Moral Psychology and Human Agency : Essays on the New Science of Ethics*. New York, NY : Oxford University Press.
- Tappolet, Christine (2016). *Emotions, Value, and Agency*. Oxford : Oxford University Press UK.
- Silva, Laura (2021a). « Anger and its desires ». *European Journal of Philosophy* 29 (4) :1115-1135.
– (2021b). « The Epistemic Role of Outlaw Emotions ». *Ergo* 8 (23).