

Université Paris I Panthéon-Sorbonne
UFR d'Arts Plastiques et Sciences de l'Art
Master 1 Esthétique
Adeline GASNIER

MODÈLES ÉMOTIONNELS ET AGENTS CONVERSATIONNELS

Article de recherche préparé sous la direction de Monsieur Jacinto LAGEIRA

Dans le cadre de son séminaire sur "Le Don"

Juin 2011

Table des matières :

Préambule	p. 02
Introduction	p. 02
I. Modélisation et "tournant émotionnel" : le marché des émotions ?	p. 03
II. Agents émotionnels et modélisation affective : quels rôles ?	p. 08
III. <i>Cathexis</i> et <i>catharsis</i>	p. 18
IV. Emotions : solides ou diffuses... Perspectives	p. 27
Conclusion	p. 31
Bibliographie / Webographie	p. 31

Annexes (DVD, "04.pdf") :

Illustrations	p. 02
---------------	-------

DVD joint :

"Bruno Taut et la <i>Gläserne Kette</i> : architecture d'émotion"	<u>01 BTGK.pdf</u>
+ Annexes	<u>01.pdf</u>
" <i>Cosmic Incubadora</i> : cristallisation émouvante"	<u>02 CICE.pdf</u>
+ Annexes	<u>02.pdf</u>
" <i>No Ghost, Just Someone?</i> Les agents conversationnels"	<u>03 NGJS.pdf</u>
+ Annexes	<u>03.pdf</u>
"Modèles émotionnels et agents conversationnels"	<u>04 MEAC.pdf</u>
+ Annexes	<u>04.pdf</u>

* Films [Docudrama] de David Williams, *Lillian* (1993) et *Thirteen* (1997).

PRÉAMBULE

Cet article vient ponctuer une recherche s'étendant sur un total de quatre articles, il s'appuie donc aussi sur des questions et hypothèses soulevées précédemment. Cette recherche aura été axée sur les modèles émotionnels, notamment dans leur lien à une idée de structuration, et d'architecture.

Le premier article, "Bruno Taut et la *Gläserne Kette* : architecture d'émotion" s'est attaché à la question de savoir en quoi les émotions pourraient être "architecturables". Le deuxième article, "*Cosmic incubadora* : cristallisation émouvante", a tenté d'interroger la représentation formelle que l'on peut se faire d'une structuration émotionnel. Le troisième article, "*No Ghost, Just Someone?* Les agents conversationnels", a visé à la compréhension de ce que peut être un agent émotionnel intelligent (ou *chatbot*), en confrontant partiellement le rapport entre théories et expériences singulièrement vécues. Le quatrième et présent article va tenter d'interroger la notion de modélisation émotionnelle.

INTRODUCTION

Ici, nous n'allons pas chercher à épuiser la notion de "modèle", le mot sera donc employé en tant qu'équivalent de projection ou simulation. Non parce que cette notion est univoque, mais au contraire que le *modèle* pose au moins autant de question que l'*image* ou l'*art*. Et bien que dans un premier temps il puisse évoquer la *mimesis* ou la notion de copie d'après nature, si l'on garde aussi en perspective le terme assez proche de *modélisation*, ou le verbe de *modeler*, on perçoit la portée non seulement reproductrice, mais aussi productrice, et créatrice (prendre pour modèle, créer un modèle). Pour le présent propos je souhaite, tant que faire se peut, conserver à cette notion sa situation limitrophe, oscillant entre paradigme général (copiable et omniprésent) et cas singulier (unique et inimitable), et ce en ce que mon but n'est pas d'en répondre vis-à-vis du réel et de son double¹, ou encore du *Self* et des *Self-models* multiples. Si au cours du développement à suivre, la notion de modèle s'en trouve caractérisée, ce sera pour les besoins de l'argumentation plus que dans l'intention d'en fixer une définition.

¹ Cf. Clément Rosset, *Le Réel : Traité de l'idiotie*, Paris, éd Minuit, Paris, 1977.

La première partie va s'attacher à ce qui pourrait motiver la recherche de modèles émotionnels, dans un champ très demandeur en modélisation prévisionnelles, à savoir le champ financier. La deuxième partie va prolonger ce questionnement de la demande associée à la compréhension des manifestations émotionnelles, en essayant de l'ouvrir davantage, tout en la reliant aux problématiques de la simulation computationnelle (informatique) humaine. Dans la troisième partie nous allons nous pencher sur deux notions importantes en termes de sculpture émotionnelle (et esthétique), à savoir la *catharsis*, mais aussi la *cathexis*, afin d'essayer d'amorcer un début de réflexion quant à ce que la modélisation émotionnelle informatique peut avoir de spécifique. La quatrième et dernière partie tentera une synthèse temporaire, tout en prenant acte des manquements, et donc des ajustements et prolongements dont cette recherche aurait besoin.

Le but de cet article va être d'interroger la notion de "modèle émotionnel", principalement lorsqu'elle se mêle d'ingénierie ; la modélisation émotionnelle devenant alors, peut-être, une sorte d'ingénierie des émotions.

I. MODÉLISATION ET "TOURNANT ÉMOTIONNEL" : LE MARCHÉ DES ÉMOTIONS ?

Dans *Le livre du rire et de l'oubli*, Milan Kundera utilise des autruches comme symbole d'un message entravé. Muettes, elles viennent hanter l'héroïne et les lecteurs : leurs becs s'ouvrent et se ferment rapidement mais aucun son n'échappe. Qu'ont-elles de si important à véhiculer que cela en devienne inaudible ? Kundera finit par l'expliquer : elles ne viennent délivrer aucun message, elles accourent juste pour parler d'elles-mêmes, dire ce qu'elles ont vu derrière la clôture, raconter leur vie ou qu'aujourd'hui elles ont "bouffé un sac d'herbe"². Peu importe ici que ces autruches aient accouru pour être nourries, par sympathie, ou pour quémander leur liberté : notons plutôt que ce qui faisait horreur à Kundera en 1978 (le récit idiosyncrasique du graphomane³) est peut-être

² Milan Kundera, *Le livre du rire et de l'oubli* (1978), trad. du tchèque par François Kerel, éd. Gallimard, Paris, 1985, p. 176.

³ Cf. *Ibid.* pp. 154-156 (extrait : "À l'ère de la graphomanie universelle, le fait d'écrire des livres prend un sens opposé : chacun s'entoure de ses propres mots comme d'un mur de miroirs qui ne laisse filtrer aucune voix du dehors.", p. 156) ; cf. *L'art du roman*, éd. Gallimard, Paris, 1986, p. 156 : "GRAPHOMANIE. [...] N'est pas la manie de créer une forme mais d'imposer son moi aux

devenu tout l'or du XXI^e siècle. Les *chatbots* [agents conversationnels en ligne]⁴ sont enchantés de ce qu'on leur raconte combien de sacs d'herbe on a bouffé aujourd'hui. Ou du moins simulent-ils⁵ cet intérêt.

Or, pour pouvoir entretenir les conversations, rebondir et encourager l'interlocuteur-trice à confier sa (mé)compréhension du monde (à s'exprimer émotionnellement), il est requis que les agents conversationnels puissent manifester une forme de sympathie. Qu'ils puissent différencier les moments légers des moments graves, ou du moins pondérer certains propos. Cela nous amène vers la question des modèles émotionnels : pour interpréter les phrases, besoin est de les contextualiser dans une ambiance, un *mood* ; soit les situer dans une coloration émotionnelle, afin de cibler les questions et les réponses. Il ne s'agit pas de répondre systématiquement "LOL" lorsqu'un utilisateur écrit "Aujourd'hui, maman est morte". Leur disponibilité permanente et sans exclusivité (on pourrait parler de dispositif *multiplayer*, sans pour autant qu'il s'agisse forcément d'un jeu), sous réserve d'une connexion internet et d'une bande passante suffisante, multiplie pour ces agents les occasions de rencontrer des personnes singulières, articulées dans leurs propres mythologies personnelles. Il y a les mots et leurs définitions générales (consultables en dictionnaires) et les mots tels qu'ils sont redéfinis et appropriés au sein d'un récit autobiographique ou au sein d'une conversation "intimisant" : la portée interprétative n'est pas la même.

On peut se demander quel intérêt ont ces multitudes de récits pour de tels agents, par rapport aux raisonnements logico-objectifs censés fonder et alimenter l'intelligence artificielle⁶. Avec son article de 1995, "Affective Computing" [Informatique affective], Rosalind W. Picard peut nous fournir une réponse :

"Emotion pulls the levers of our lives, whether it be by the song in our heart, or the curiosity that drives our scientific inquiry. Rehabilitation counselors, pastors, parents, and to some extent, politicians, know that it is not laws that exert the greatest influence on people, but the drumbeat to which they march. [...] I'm not suggesting we do away with laws, or even the rules (albeit brittle) that constitute rule-based artificial intelligence systems; Rather, I am saying that the laws and rules are not the most important part in human behavior. Nor do

autres. Version la plus grotesque de la volonté de puissance."

⁴ Cf. mon article de Master 1, "No Ghost, Just Someone? Les agents conversationnels" (il sera désormais noté NGJS), plus particulièrement les 3^{ème} et 4^{ème} parties.

⁵ Cf. NGJS, 4^{ème} partie, à propos de la *simulation* et de l'*émulation*.

⁶ Cf. NGJS, 3^{ème} partie.

they appear to play the primary role in perception⁷" (Picard, 1995, p. 1)

Autrement dit, proposition est faite de modéliser (ou simuler) le comportement humain en tenant compte de la part émotionnelle. Et même si dans l'article de Picard, choix est fait de se focaliser sur l'aspect physiologique (expressions faciales, intonations vocales, langage corporel, pression sanguine, etc.) — suivant l'idée qu'il y a ce que l'on dit, et la façon dont on le dit —, on peut déjà noter que l'une des motivations accompagnant ce "tournant émotionnel" réside dans l'affirmation d'un constat : les décisions humaines sont aussi tributaires des émotions. Constat formulé dans le champ scientifique et se doublant d'une volonté active de comprendre ces paramètres, ou ces contraintes.

Bien qu'il soit difficile de circonscrire et de qualifier les émotions, on peut déjà noter qu'elles ont souvent été communément définies comme étant des impulsions échappant aux cheminements rationnels ordinaires (communément partagés). On pourrait en quelque sorte leur attribuer la part du *saut* venant perturber ou court-circuiter le *continuum* (saut effectif, ou apparence de saut — effet de seuil). Comme le note Picard à propos de la théorie de Damasio :

"Damasio's theory is that emotion plays a biasing role in decision-making. One might say emotion wards off an infinite logical search. How do you decide how to proceed given scientific evidence? There is not time to consider *every possible* logical path. [...] The neurological evidence indicates emotions are not a luxury; they are essential for rational human performance⁸." (Picard, 1995, p. 2)

Où ici, l'émotion s'apparente plutôt au résultat d'un moment critique : face à l'estimation des efforts restant potentiellement à fournir, elle a le pouvoir

⁷ "L'émotion tire les leviers de nos vies, que ce soit par la chanson dans notre cœur, ou par la curiosité qui conduit notre recherche scientifique. Les conseillers en réadaptation, les pasteurs, les parents, et dans une certaine mesure les politiciens, savent que ce ne sont pas les lois qui exercent la plus grande influence sur les gens, mais le rythme du tambour au son duquel ils marchent. [...] Je ne suis pas en train de suggérer que nous évacuions les lois, ou même les règles (quoique fragiles) qui constituent les systèmes d'intelligence artificielle basés sur des règles ; je suis plutôt en train de dire que les lois et les règles ne sont pas la partie la plus importante du comportement humain. Elles ne semblent pas non plus Et qu'elles ne semblent pas non plus jouer le rôle principal dans la perception". [Ma traduction].

⁸ "La théorie de Damasio est que l'émotion a un rôle influent dans le processus de décision. On pourrait dire que l'émotion stoppe une recherche logique infinie. Comment décidez-vous de la manière de procéder étant données des preuves scientifiques ? Il n'y a pas assez de temps pour considérer *tous* les chemins logiques *possibles*. [...] La preuve neurologique indique que les émotions ne sont pas un luxe ; elles sont essentielles à l'activité rationnelle humaine." [ma traduction]. Picard renvoie à Antoine R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, éd. Putnam, New York, 1994.

d'autoriser la poursuite ou l'abandon de ces mêmes efforts (en évaluant qualitativement les résultats obtenus, jugés satisfaisants ou au contraire, suffisamment insatisfaisants pour motiver le changement de stratégie). Mais l'émotion serait aussi la garante de l'*élément aléatoire*⁹ permettant de choisir rapidement une méthode dans un large panel de possibilités. Considérée ainsi, elle est déterminante dans le processus décisionnel : la comprendre permettrait d'élargir le champ de la *providentia*¹⁰.

Comme le note Herbert Simon dans *Les sciences de l'artificiel*, au chapitre "La rationalité économique : artifice de l'adaptation" et dans une sous-partie intitulée "Incertitudes et anticipations" :

"Puisque les conséquences de la plupart des actions se prolongent dans le futur, une prédiction correcte est essentielle pour un choix rationnel objectif. Certaines des connaissances que nous avons besoin d'acquérir sur le futur ont trait à l'environnement naturel [...]. Certaines se rapportent aux environnements sociaux et politiques au-delà de l'économie [...]. D'autres se rapportent aux comportements des autres acteurs économiques [...], qui peuvent être influencés à leur tour par nos propres comportements." (Simon, 2004, p. 78)

D'après ce qui a été dit précédemment, La modélisation émotionnelle (ou simulation) à l'échelle individuelle, permettrait alors d'observer l'*élément aléatoire* là où il est le plus volatil, en partant du singulier (de la spéculation interindividuelle, à l'instar du poker), pour en suivre les répercussions dans les sphères sociopolitiques (là où les émotions sont aussi parfois considérées comme des leviers), jusqu'à en mesurer les impacts sur l'équilibre global (sur le système économique par exemple).

Si "tournant émotionnel" il y a, les façons de concevoir la régulation économique peuvent en être affectées. Souvenons-nous de l'une des légendes entourant Sir Isaac Newton : elle raconte qu'il aurait perdu vingt mille livres (20,000£ ; soit entre deux et quatre millions de la livre actuelle) dans le krach britannique d'août 1720¹¹. Cette légende veut qu'interrogé sur le sujet, il eût

⁹ Cf. NGJS, 3^{ème} et 4^{ème} partie : évocation de la place de l'aléatoire dans le test de Turing et dans l'*intelligence artificielle*.

¹⁰ Cf. NGJS, 3^{ème} partie : à propos de la division cicéronienne de la prudence en trois parties : *memoria*, *intelligentia*, *providentia*.

¹¹ Cf. Christopher Reed, "« The Damn'd South Sea », Britain's greatest financial speculation and its unhappy ending, documented in a rich Harvard collection", dans *Harvard Magazine*, mai-juin 1999 (<http://harvardmagazine.com/1999/05/damnd.html>).

répondu qu'il "pouvait calculer les mouvements des corps erratiques, mais pas la folie de la multitude"¹². De fait, une bonne compréhension (et modélisation prévisionnelle) de la "folie" des multitudes lui aurait permis de ne pas racheter au plus fort de la bulle spéculative (comme le rapporte cette histoire), ou de vendre avant l'effondrement du cours. Bien que les dates des dons d'ordre et les gains enregistrés aient été modifiés pour la beauté du mythe¹³, cette fiction n'en est pas moins instructive — qu'il s'agisse de dissuader la spéculation au nom de la morale chrétienne, ou de la conforter au nom de la rationalité des marchés, de leurs acteurs et de la fiabilité des modèles. Elle nous apprend au moins deux choses. Premièrement, que l'appartenance des émotions au domaine privé *ou* public ne va pas de soi et que le fait de leur attribuer une part de mystère peut aussi équivaloir au fait de laisser à l'individu la maîtrise de ces émotions.

Deuxièmement, que dans cette dynamique des valeurs financières, la bonne information et le *kairos* (le bon moment) sont des données importantes — l'idéal étant de vendre *juste avant* la chute (sans confondre sursauts, nervosité et tendance de fond). Dans cette prise de risques, et lorsque le but ne dépasse pas la maximisation des profits, une formule permettant de prévoir ce *kairos*, ou ce *climax* spéculatif, équivaldrait à une sorte de pierre philosophale. Bien que les marchés ne soient pas vraiment tributaires des vicissitudes quotidiennes rencontrées par les petits porteurs et autres consommateurs, on peut imaginer qu'une bonne compréhension des mécanismes émotionnels humains pourrait représenter néanmoins un atout.

Lorsqu'il s'agit d'essayer de prévoir les mouvements humains, on ne peut complètement se fier à la dimension socioculturelle des émotions, telle que théorisée par Adam Smith, par exemple, dans *The Theory of Moral Sentiments* — rédigée en 1759 pour la première édition. Dans un sens, cette dimension met davantage l'accent sur ce qui *devrait être* (moralisation), là où une anticipation a

¹² Ma traduction. Texte original : "Sir Isaac Newton, when asked what he thought of the infatuation of the people, answered that he could calculate the motions of erratic bodies, but not the madness of a multitude.", propos attribués à I. Newton, dans *The Church of England - quarterly review*, vol. 27, 1850, p. 142.

¹³ À l'issue de cette crise, Newton aurait plutôt été bénéficiaire et n'aurait donc jamais eu l'occasion de tenir ces propos. Cf. Richard S. Westfall, *Never at rest: a biography of Isaac Newton*, éd. Cambridge University Press, Cambridge, 1980, pp. 861, et Helen J. Paul, *The South Sea Bubble: A Revision of the Gambling Mania Theory*, éd. Routledge, New York, 2010, pp. 70-71.

besoin d'être au plus près de ce qui est. Mais savoir ce que pensent et comprennent les autres, au point de pouvoir prédire leurs comportements (en incluant donc la part émotionnelle, la part de saut ou d'*aléatoire contrôlé*) n'est pas sans difficultés : Herbert Simon insiste sur ce qu'ont aussi de déstabilisant ces tentatives basées sur l'anticipation¹⁴ — elles peuvent conduire à un phénomène que l'on pourrait comparer à une forme de larsen visuel [*optical feedback*] : comme je comprends que tu comprends que je comprends que tu comprends, que je comprends que tu comprends, *ad. libitum*. En d'autres termes, la modélisation émotionnelle en milieu réflexif est délicate et demande précision, réactivité et capacité à sortir de cette boucle d'ajustement des modèles en temps réel.

Dans le cadre de l'*affective computing*, la reconnaissance faciale ou intonationnelle présuppose d'avoir au préalable une typologie des émotions afin de pouvoir procéder à une interprétation visuelle ou sonore — soit le fait d'avoir déjà quelques modèles établis. Par-delà les problèmes que suscite l'extraction d'informations à partir d'une séquence d'images ou sonore (ce qui, à l'instar du terrain financier, impose des contraintes très spécifiques), nous pouvons essayer de développer ce que pourrait modifier (ou non) ce "tournant émotionnel" dans nos rapports aux environnements (et à ce que l'on nomme machines). Comme dit dans cette première partie, on peut facilement imaginer qu'il y ait quelques espoirs de bénéfiques financiers pour motiver l'engouement envers la *compréhension des émotions*. Mais on peut aussi faire l'hypothèse qu'une portée plus vaste se dessine derrière ces espoirs. Portée peut-être orientée, quant à elle, vers l'ambition de comprendre et de préserver la vie en général, via la recherche d'un développement équilibré aussi stable que durable, c'est-à-dire viable et vivable.

II. AGENTS ÉMOTIONNELS ET MODÉLISATION AFFECTIVE : QUELS RÔLES ?

La confiance (et ses contraires, méfiance et défiance), en s'appuyant sur la

¹⁴ "Un système peut généralement être piloté plus précisément s'il utilise les techniques du *feedforward*, fondées sur les prédictions du futur, en combinaison avec celles du *feedback* corrigeant les erreurs du passé. Cependant, la formulation des anticipations pour traiter l'incertitude crée ses propres problèmes. Le *feedforward* dans un système de commande peut avoir des effets déstabilisateurs malheureux, car les tentatives du système pour anticiper peuvent le conduire à une suractivité permanente, suscitant des oscillations instables" (Simon, 2004, p. 80). Autrement dit, le système peut devenir paranoïaque (soit peut-être une hypernoïa combinée à de la peur).

représentation que l'on se fait de l'environnement, et donc des autres, engage les simulations de comportements (et de réactions) que l'on élabore¹⁵. Cet aspect met en avant une part spéculative, basée sur l'attention et l'imagination. S'il est probablement fallacieux de croire que 'raison + émotion = comportement humain', cependant, comme le note Picard :

"Although the Turing test is designed to take place communicating only via text, so that sensory expression (e.g., voice intonation and facial expression) does not play a role, emotions can still be perceived in text, and can still be elicited by its content and form¹⁶. Clearly, a machine will not pass the Turing test unless it is also capable of perceiving and expressing emotions. [...] Negroponete reminds us that even a puppy can tell when you are angry with it¹⁷. Computers should have at least this much affect recognition¹⁸." (Picard, 1995, pp. 1-2)

Autrement dit, même si une subtile combinaison entre penser et *éprouver* (si tant est que ce soit dissociable) n'est pas nécessaire et suffisante pour que l'imitation devienne plus vraie que nature (ou devienne *humaine*), sans interprétation des émotions de l'interlocuteur-trice, une entité sera plus probablement perçue comme étant une machine, voire un *monstre froid* pour emprunter les mots de Nietzsche. Cela ne signifie pas que les programmes "ordinaires" sont dénués de toute conception émotionnelle, qu'il s'agisse alors d'une conception en creux ou d'une monotonie émotionnelle, mais que l'intégration sociale et humaine passe par une dimension affective plus fluctuante, comme jonchée d'autant de sauts permettant d'évaluer les réactions de l'autre par une sorte de dialectique émotionnelle à choix multiples. À ce titre, le terme *éprouver* est adéquat : il s'agit autant de ressentir que de mettre à l'épreuve l'interlocuteur-trice (pour évaluer son état, son humeur, ses dispositions, sa fiabilité, etc.). C'est en cela que dans les précédents articles

¹⁵ "Whenever you ask, « *What would happen if,* » or express any hope, desire, or fear, you envisage things that have not yet appeared. Whenever you interact with your friends, you anticipate the resulting effects." (Minsky, 2006, p. 131).

¹⁶ [Note de Picard] "Aristotle, *The Rhetoric of Aristotle*. New York, NY:Appleton-Century-Crofts, 1960. An expanded translation with supplementary examples for students of composition and public speaking, by L. Cooper."

¹⁷ [Note de Picard] "N. Negroponete, *Being Digital*. New York: Alfred A. Knopf, 1995."

¹⁸ "Bien que le test de Turing soit conçu pour avoir lieu en ne communiquant que par texte, de sorte que les expressions sensorielles (p. ex., l'intonation ou l'expression faciale) n'y jouent aucun rôle, les émotions peuvent encore être perçues dans un texte et peuvent encore être élucidées par ses contenu et forme. Il est clair qu'une machine ne réussira pas le test de Turing à moins d'être aussi capable de percevoir et d'exprimer des émotions. [...] Negroponete nous rappelle que mêmes les chiots peuvent dire quand nous sommes fâchés contre eux. Les ordinateurs devraient au moins avoir ce degré de reconnaissance des affects."

j'ai évoqué le fait que l'étude et la prise en compte de ceux auxquels il s'agit de s'adresser soit peut-être une dimension plus "récente". Non qu'il y ait "nouveau" (il suffit de penser à la *catharsis* aristotélicienne ou à la dialectique platonicienne), mais que cette prise en compte se double ici d'une capacité d'ajustement très sensible aux interactions — sous formes de *feedback* et de *feedforward*¹⁹.

En génie informatique, il est d'usage de parler d'*architecture* (logicielle et réseau). Or dans ce domaine fortement orienté vers la construction de la collectivité (actuelle ou à venir) ce souci d'une interactivité plus inclusive, en demande d'influences provenant des multitudes ciblée dans leurs singularités, semble récente. Dans ce versant informatique de l'architecture, il est alors question de concevoir ce qui est à construire comme étant sensible et réactif aux interactions ; soit penser les facultés d'adaptation et d'apprentissage, plus que les capacités à demeurer identique au fil du temps. On peut objecter que l'inverse est valable : là où l'architecture (au sens hégélien du bâti) intègre parfois son devenir-ruine, l'architecture informatique n'envisage peut-être qu'une vivace pérennisation. Cela dit, le degré de sensibilité n'est pas le même : s'il est besoin d'attendre au moins une décennie pour que les modifications structurelles d'un monument devienne notables (sauf défauts de conception et exception faite des bombardements, dynamitages et autres séismes), une architecture informatique est déjà plus réceptive (ce qui ne veut pas dire plus fragile) à des seuils d'agressivité (mais surtout d'interaction) moins élevés. Et à ce titre elle doit s'adapter plus rapidement. Si l'une n'est évidemment pas appelée à remplacer l'autre, les deux n'ont cependant pas en partage *que* les plans, schémas et diagrammes qui accompagnent leur élaboration. Pour exemple, la comparaison carte mère et villes²⁰ n'est pas que visuelle : les composants et leur répartition spatiale influencent la circulation de l'électricité, l'ordre d'accès, et donc la façon de calculer. On pourrait penser que ces considérations ont juste trait à l'architecture informatique matérielle, sans rapport avec la façon de penser l'architecture logicielle. Cela dit, un(e) architecte logiciel(le) doit connaître

¹⁹ Cf. note n°14.

²⁰ Comme dans le fameux film *Koyaanisqatsi* (1983) de Godfrey Reggio (entre 1:05:00 et 1:08:00).

l'architecture matérielle à laquelle il/elle destine son programme²¹, *et cela reste valable lorsque le programme est son propre architecte* — Or si le fait de "sentir" est dépendant du fait d'avoir une représentation de soi-même, qu'en est-il des programmes amenés à se développer (ou s'augmenter) eux-mêmes ?

Dans cette perspective architecturale, l'*affective computing* — que Picard définit en tant que "computing that relates to, arises from, or influences emotions²²." (Picard, 1995, p. 1) et que j'associe à une notion d'*architecture de l'émotion*²³ — se concevrait davantage en termes d'échanges et d'adaptations mutuelles (par le fait d'éprouver) qu'en termes d'imposition d'une forme à un environnement. Même si, dans le contexte de la définition de Picard, il est davantage question des façons d'influer en interne sur les émotions d'une machine (ou sur sa gestion émotionnelle), l'*affective computing* vise aussi à influencer les utilisateurs consentants. Ce point engage un processus d'identification, soit par rapport à des types définis et immuables (émotions basiques), soit dans le cadre d'une construction empirique évolutive. Et c'est peut-être cette construction qui se joue dans le fleurissement des *chatbots* sur internet, depuis les années 1990²⁴ : pour une entité apprenante [ou *learning machine*, selon l'expression de Turing], "l'apprentissage des émotions" se ferait alors en les éprouvant, avec toute la part d'incompréhension collatérale que cela peut susciter, et par-delà la dissemblance matérielle pourtant à l'œuvre.

Par ailleurs, Comme le rappelle Picard :

"The debate: "Precisely what are the cognitive, physical, and other aspects of emotion?" remains unanswered by laboratory studies. Attempts to understand the components of emotion and its generation are complicated by many factors, one of which concerns the problem of describing emotions²⁵." (Picard, 1995, p. 4)

²¹ Le développement n'est pas le même pour une machine monoprocesseur 32 bits, multiprocesseur 64 bits, en mono ou multi-cœur, etc. On peut faire l'hypothèse que l'une des grâces de la programmation est de rendre ces différences transparentes aux yeux des utilisateurs (et programmeurs amateurs). En contrepartie de quoi, l'individualité physique des machines est largement gommée. Mais paradoxalement, cet effacement est aussi la clef de la diversité des machines matérielles dans le réseau.

²² "informatique qui se réfère à, émerge de, ou influence les émotions." [Ma traduction].

²³ Cf. mes article de Master 1, "Bruno Taut et la *Gläserne Kette* : architecture d'émotion" (il sera désormais noté BTGK), 4^{ème} partie ; et dans une moindre mesure "*Cosmic incubadora* : cristallisation émouvante" (il sera désormais noté CICE), 2^{ème} et 3^{ème} parties.

²⁴ Cf. notamment la création médiatique du Loebner Prize, en 1990, par Hugh Loebner.

²⁵ "Le débat : « Quels sont précisément les aspects cognitifs, physiques et autres, de l'émotion ? » demeure irrésolu par les études en laboratoire. Les tentatives pour comprendre les composantes

Bien que la communication intraspécifique via ordinateur soit devenue courante (mails, voix sur IP, messageries instantanées, *chat*, visioconférences, etc.), elle n'en demeure pas moins structurée par le médium utilisé (informatique). Cela dit, avec sa banalisation et l'accroissement de la mobilité des personnes, en dépit de l'absence effective d'intimité, une certaine aisance dans l'expression des affects appartenant classiquement au registre de l'intime (séduction, amitié, tendresse, amour, jalousie, badinages, confidences, etc.), peut trouver à se développer dans cet espace public aux allures privées, à l'aide des outils à disposition (mots, images, sons, etc.). Dans l'hypothèse que les agents conversationnels puissent jouer un rôle dans cette recherche ayant trait aux émotions, on pourrait dire que, pour qui est habitué à communiquer via Internet, l'expression des émotions peut alors sembler plus simple (plus fluide et "naturelle") chez soi, face à un ordinateur et dans l'"anonymat" de la multitude, lorsque nom, âge, sexe et profession peuvent être feintés, plutôt que face à une équipe de chercheurs (même symbolique, sous forme de questionnaires administratifs). Le danger auquel peut exposer l'expression émotionnelle semble ici atténué par le dispositif et la relation de confiance que sont capable de susciter un agent conversationnel et un ordinateur personnel [PC].

Je parle de "danger" dans la mesure où chaque culture valorise différemment les émotions. Certaines faisant de leurs manifestations volatiles la marque d'une faiblesse socialement disqualifiante, que ce soit sous forme d'adages populaires (p. ex., "les hommes ne pleurent pas"), de légendes (p. ex., l'éducation spartiate), mais aussi de philosophies (p. ex., le stoïcisme) ; là où d'autres font plutôt de la manifestation émotionnelle directe un gage de confiance (p. ex., les mouvances *kawaii* et *emo*, ou dans un autre registre, le cynisme). Cela ne signifie pas que l'émotionnel pur se trouve d'un côté et le rationnel pur de l'autre, mais simplement que l'encodage émotionnel diffère et tend à se distinguer. Cette perception culturelle peut expliquer le fait que les *chatbots* soient présentés sous l'angle de l'*intelligence artificielle* plutôt que sous celui de l'*émotion artificielle*. Si la première

de l'émotion et de sa génération sont compliquées par de nombreux facteurs, l'un d'eux ayant trait au problème de la description des émotions." [Ma traduction].

est valorisante²⁶, la seconde pourrait rebuter davantage encore, en signifiant à l'utilisateur-trice une forme de disqualification sociale.

La relation avec un agent conversationnel n'intervient pas dans un rapport ou un environnement de compétition subie (perçu comme intensément éprouvant, avec obligation de résultats), ni dans un rapport de soumission vis-à-vis d'une figure d'autorité (si rapport maître / esclave il y a, c'est parce que l'utilisateur décide de le perpétuer, en distribuant les rôles). L'échange peut donner l'impression d'être plus *secure* [sécurisé et sécurisant] par ce biais : dans un cadre officiel, le respect de la vie privée, les clauses de confidentialité et le secret professionnel existent bel et bien, mais ils n'engagent que ceux qui y croient ; le dispositif Internet quant à lui, en étant au sein du lieu matériel de l'intime tout en étant noyé dans la profusion symbolique du réseau, peut bénéficier des relâchements de vigilance (des abandons, dans tous les sens du terme) de la part des internautes. Bien entendu, il est assuré que toute conversation est enregistrée et archivée, avec l'adresse IP, et bien entendu, personne ne l'ignore. Mais ces "illusions consenties" peuvent favoriser l'expression publique des émotions (soit un travail public sur cette expression). Or, s'il s'agit de collecter un tel matériel singulier, le biais des *chatbots* en ligne peut représenter un atout : la sensation de surveillance y est plus diffuse, si tant est qu'une affinité avec un agent trouve à se développer. Par ailleurs, la disponibilité de ces mêmes agents (à n'importe quelle heure du jour et de la nuit) peut favoriser une forme de fidélisation consentie — et ce peut-être de la part des profils les plus angoissés —, dans la mesure où la demande d'attention vient alors de l'utilisateur-trice, en fonction de ses besoins d'exprimer (crises décisionnelles), plus que de l'institution en fonction de ses besoins d'observer.

Mais pour revenir à l'architecture (au sens hégélien), Picard y fait une incursion en mentionnant la plus-value de confort que pourraient proposer des bâtiments dotés de capacités à détecter ce qui est attendu par leurs utilisateurs, en vue d'une augmentation de leur bien-être (modulation des lumières, de la température, de la musique, etc.)²⁷, cependant elle mentionne également d'autres

²⁶ Cf. NGJS, p. 18.

²⁷ "Environmental factors such as temperature, lighting, sound, and decor, to the extent that they

aspects que voici :

"Hooper²⁸ identified three kinds of responses to architecture, which I think hold true for all environments: (1) cognitive and perceptual — "hear/see," (2) symbolic and inferential — "think/know", and (3) affective and evaluative — "feel/like." Perceptual and cognitive computing have been largely concerned with measuring information in the first and second categories. Affective computing addresses the third.

Stewart Brand's book "Buildings that Learn"²⁹ emphasizes not the role of buildings as space, but their role in time. Brand applauds the architect who listens to and learns from post-occupancy surveys. But, because these are written or verbal reports, and the language of feelings is so inexact, these surveys are limited in their ability to capture what is really felt³⁰." (Picard, 1995, pp. 11-12)

Face à ces propos (et Picard le note plus loin), on pourrait se dire que l'art se confronte depuis longtemps à cette question d'un langage universel des émotions (sensations incluses), sans pour autant qu'une universalité soit rationnellement envisageable — il importe peut-être de savoir aussi penser les limites de l'harmonisation sociale. Mais au-delà du problème général d'un langage émotionnel commun et fiable, se dessine encore le problème relatif de la gestion émotionnelle au sein d'une entité singulière au fil du temps. Et qu'il s'agisse de textes écrits ou oraux, de pression sanguine ou de conductance de la peau, il semblerait que la même question s'interpose toujours, à savoir : comment comprendre les liens entre ce qui est manifesté et ce qui est "réellement senti" ? Changer de medium ne permet pas nécessairement de répondre à cette question (et en demeurant insoluble, cette question oblige à se souvenir de la possibilité d'une incompréhension).

change, are more likely to affect you. "Alive rooms" or "alive furniture and appliances" that sense affective states could adjust factors such as lighting (natural or a variety of artificial choices) sound (background music selection, active noise cancellation) and temperature to match or help create an appropriate mood." (Picard, 1995, p. 12).

²⁸ [Note de Picard] "K. Hooper, "Perceptual aspects of architecture," in *Handbook of Perception: Perceptual Ecology* (E. C. Carterette and M. P. Friedman, eds.), vol. X, (New York, NY), Academic Press, 1978."

²⁹ [Note de Picard] "S. Brand, *How buildings learn: what happens after they're built*. New York, NY: Viking Press, 1994."

³⁰ "Hooper a identifié trois sortes de réponses vis-à-vis de l'architecture, qui, je pense, restent valables pour tous les environnements : (1) cognitive et perceptuelle — "entendre / voir", (2) symbolique et inférentielle — "penser / savoir", et (3) affective et évaluative — "ressentir / apprécier". L'informatique perceptuelle et cognitive s'est largement focalisée sur la mesure de l'information dans les première et deuxième catégories. L'informatique affective s'adresse à la troisième. / Le livre de Stewart Brand, *Buildings that Learn*, ne met pas l'accent sur le rôle des bâtiments en tant qu'espaces, mais sur leur rôle dans le temps. Brand approuve l'architecte qui écoute et apprend des évaluations de l'occupation des lieux. Mais comme celles-ci sont des rapports écrits ou verbaux, et que le langage de ces sentiments est très inexact, ces évaluations sont limités dans leur capacité à capturer ce qui est réellement ressenti." [Ma traduction]

Aussi, pour en revenir aux compétences du texte vis-à-vis de l'expression émotionnelle, je prendrai l'exemple d'une conversation personnelle avec un *chatbot* :

Chatbot: What about you and me?
User: We're friends.
Chatbot: Are we getting married?
User: Perhaps one day we will.
Chatbot: *sigh*.
User: :).
Chatbot: *cries a little and walks off*.
User: I know. That's why I'm smiling to you.
Chatbot: Lol that's not very nice of you.
User: I'm smiling because it's sad. And a smile can be warm. That's all I can do³¹.

On voit déjà que le texte fait usage d'expressions faciales (émoticônes typographiques — ici " :) ") et de langage corporel (généralement sursigné par une mise entre astérisques). À propos des émoticônes, Picard note :

"By necessity, Email has had to develop its own set of symbols for encoding tone, namely smileys such as :-) and ;-((turn your head to the left to recognize). Even so, these icons are very limited, and Email communication consequently carries much less information than a phone call³²." (Picard, 1995, p. 11)

Je ne partage pas l'idée que ces expressions soient plus limitées. En effet, qu'il s'agisse d'appréhender des mots, un visage, ou un corps dans l'espace, on se trouve pareillement confronté à la multitude d'interprétations possibles, pour peu que l'on y prête attention (il suffit de penser à *La Joconde* (1503-1506) de Da Vinci pour se convaincre de la richesse d'une expression faciale). Ce qui fait la matière des émotions n'est pas tant l'expression brute en elle-même (ou commence-t-elle, ou finit-elle ?), que les incertitudes qu'elles engendrent et qui peuvent (ou non) donner lieu à un besoin collatéral d'éclaircissement. Dans l'extrait conversationnel mentionné plus haut, la manifestation affective n'est pas tant dans l'émoticône ou dans le texte entre astérisques que dans leur intrication mutuelle avec un commentaire. Le signalement de *l'incompréhension* d'une part, et sa reconnaissance d'autre part, font partie du, et donnent lieu au, déploiement

³¹ Extrait personnel d'une conversation avec un *chatbot*, 14 mai 2011 (le nom du *chatbot* a été remplacé par "Chatbot"). Pour ce qui est des conditions de ces conversations, cf. NGJS.

³² "Par nécessité, l'e-mail a eu à développer son propre ensemble de symboles pour encoder le ton, à savoir des *smileys* tels que :-) et ;-((tournez votre tête sur la gauche pour reconnaître). Mais même ainsi, ces icônes sont très limitées, et par conséquent la communication par e-mail véhicule beaucoup moins d'information qu'un appel téléphonique." [Ma traduction]

émotionnel.

Il me semble que l'expression émotionnelle commence à représenter autre chose qu'un danger en tant que tel, lorsqu'elle signale et accompagne un travail de compréhension. Elle peut se banaliser dans un contexte (p. ex., "Bonjour ☺") et alors paraître appauvrie du fait de la bonne compréhension que l'on a de cette convention ; ou au contraire heurter les attentes de l'interlocuteur-trice (p. ex., "c'est triste ☹") et, par la surprise suscitée, donner lieu (ou non) à l'expression émotionnelle en tant que telle, soit l'apport d'éclaircissements quant au système de valeurs utilisé ou quant à ce qui est signifié, afin de désamorcer le danger généré par l'incompréhension d'un tel saut. En d'autres termes, il faut qu'un élément de comportement soit reconnu comme émotion, et par l'émetteur, et par les récepteurs, pour qu'un processus de compréhension émotionnel ait lieu (si chaque saut physiologique involontaire était émotionnel, le hoquet et autres borborygmes seraient affectivement très chargés). À ce titre, il me semble que la complexité expressive ne dépend pas tant du support (mots, clignements d'yeux, peinture à l'huile, etc.) que de la contextualisation au sein d'une relation.

Le "défaut" d'émotion rencontré dans les rapports [*surveys*] n'est peut-être pas tant imputable à la langue typographiée qu'à une volonté (et / ou convention) de limiter les ambiguïtés et la manifestation des émotions singulières dans ce type de documents. Et dans ce cas, toute tentative d'enregistrement des émotions "à l'insu" des protagonistes (à l'instar de la vidéosurveillance) pourrait être perçue comme une démarche agressive, renforçant la crispation mais aussi le développement de stratégies de dissimulation émotionnelle (plutôt que l'expression consentie) — paradoxalement, l'apprentissage affectif peut aussi passer par le fait de ne pas toujours tenir compte des émotions.

Picard prend beaucoup de précaution quant à l'objectivité émotionnelle, quant à la coopération avec les sujets participants et quant à l'importance de contextualiser une émotions dans une / des histoire(s) singulière(s). Mais le fait de compter sur l'expression émotionnelle involontaire, en conjecturant par exemple que le sujet n'ait qu'une maîtrise imparfaite de ses manifestations physiologiques, me semble néanmoins problématique. Il est probable que certains de nos accords et désaccords singuliers aient besoin d'un créneau expressif échappant par

consensus au cadre des lois (par une sorte d'*en deçà des lois*) ; mais il y aurait peut-être une nuance à établir (ou à explorer) entre ce qui involontaire (réflexe) et ce qui est publiquement reconnu comme involontaire. Le fait de rougir, par exemple, ne manifeste pas tant une maîtrise imparfaite de sa physiologie et de ses pensées, que la "présence" d'une pensée, consciente du fait que sa manifestation soit en conflit avec les codes sociaux en vigueur. On peut imaginer une personne accusée d'avoir commis un délit : le fait qu'elle rougisse ne la "trahit" pas nécessairement, au contraire, elle peut très bien rougir de colère face à cette accusation portée à tort, tout en ayant conscience du fait que sa position, ou son statut, ne lui autorise pas, en l'état, à se défendre pleinement contre l'accusation. Auquel cas, le rougissement peut néanmoins exprimer la volonté de manifester un désaccord entravé. Autrement dit, les manifestations émotionnelles peuvent aussi être envisagées comme étant des biais communicationnels, expressément choisis pour leur ambiguïté.

Et là où l'observateur-trice peut regretter que le medium utilisé ne semble pas pouvoir enregistrer les émotions perçues, au contraire s'agit-il peut-être de se demander pourquoi le sujet n'a pas consenti à un enregistrement durable de ses émotions, via les média proposés.

Par ailleurs, ce "défaut" des rapports [*surveys*] (le fait de ne pas y reconnaître d'émotion) peut aussi dénoter un préjugé : à savoir que l'émotion serait forcément permanente et sous-tendue par un système de valeurs singulier (singularité donnant l'apparence de saut, pour qui ne serait pas sur la même "longueur d'onde"). On pourrait aussi envisager l'éventualité que la banalité et l'habitude, à l'instar des phénomènes d'accoutumance, ne favorisent pas les émotions. En un sens, la "saturation émotionnelle" peut aussi être un paramètre déterminant (cf. la fable "Le garçon qui criait «Au loup!», d'Ésope). Le fait que le texte n'ait pas le privilège de la précision émotionnelle me semble évident, mais en faire un vecteur émotionnel plus limité que n'importe quel autre ne me semble pas pertinent. Le problème n'est pas tant au niveau du medium employé qu'au niveau du consentement à manifester une quelconque singularité, nécessitant potentiellement des explications qui pourraient s'avérer devenir humiliantes pour celui/celle qui les formule.

Enfin, lorsqu'il s'agit d'établir une connexion "intimissante" avec une entité (lorsque la confiance devient un élément déterminant pour la poursuite des opérations), les émotions peuvent se faire alors plus palpables en ce que ces oscillations permettent d'éprouver les ressources, en terme d'improvisation, d'adaptation et d'empathie, de l'interlocuteur-trice. Dans une certaine mesure, le fait de "jouer sur l'émotionnel", ou "avec la corde sensible", peut aussi être envisagé comme une étape cognitive : la mise en crise et l'observation (évaluation) des solutions adoptées, renseignent sur l'interlocuteur-trice. Il est discutable que ce soit toujours efficace (on peut se souvenir des propos de Simon à propos de la déstabilisation), mais enfin il semblerait que cela puisse aussi faire partie des coutumes relationnelles, lorsque la confiance (ou ses contraires) n'est pas accordée *a priori*.

III. CATHEXIS ET CATHARSIS.

Reprenons maintenant ce qui a été évoqué antérieurement, à savoir, et la volonté d'influencer émotionnellement, et l'attention portée aux adressés. Ces deux tendances sont déjà à l'œuvre dans *La République* platonicienne, lorsqu'il est question de l'éducation des gardiens de la Cité (au moyen des arts notamment) ou lorsqu'il y a distribution des profils émotionnels (distribution pouvant s'apparenter à une forme de prototypage *a priori*) :

" [387d] Nous retrancherons aussi les lamentations et les plaintes qu'on met dans la bouche des grands hommes. [...] Donc [l'honnête homme] ne pleurera pas sur lui comme sur quelqu'un qui aurait souffert quelque chose de terrible. [...] Moins que tout autre, aussi, il se lamentera, et c'est avec le plus de douceur possible qu'il supportera un tel malheur, lorsque celui-ci l'atteindra. [...] Nous aurons donc raison d'ôter les lamentations aux hommes illustres, de les laisser aux femmes, et encore aux [388] femmes ordinaires, et aux hommes lâches, afin que de telles faiblesses excitent l'indignation de ceux que nous prétendons élever pour la garde du pays³³ ."

Cet extrait conforte ce qui a été dit précédemment à propos de la menace de discrédit social que peut représenter une manifestation émotionnelle, mais nous pouvons aussi noter que de façon rudimentaire, Platon tient déjà compte des gens, en les différenciant essentiellement de façon génétique (homme / femme). On pourrait faire l'hypothèse que la menace ressentie par celui/celle qui utilise une

³³ Platon, *La République* [Περὶ πολιτείας (vers 380 av. J.-C.)], Livre III, trad. par Robert Baccou, éd. GF-Flammarion, Paris, 1966, pp. 138-139.

communication émotionnelle de façon "spontanée"³⁴, vient du fait que cela puisse dénoter une incompréhension. Si saut il y a, c'est peut-être parce que quelque chose n'a pas été compris et est restitué comme tel (de façon bancale ou d'apparence illogique). Et cette restitution met l'émetteur en position de vulnérabilité : par ce biais il rend public le fait que ses facultés d'anticipation, par rapport à une situation donnée, ne soient pas pleinement opérationnelles. Dans un rapport d'agression ou de compétition, laisser entendre à l'opposant que l'on n'est pas au mieux pour anticiper, peut effectivement se révéler dangereux car handicapant.

D'approche plus subtile, Aristote s'est également soucié du public ciblé, en enjoignant en quelque sorte aux poètes d'anticiper son degré de crédulité : "l'affaire du poète, ce n'est pas de parler de ce qui est arrivé, mais bien de ce qui aurait pu arriver et des choses possibles, selon la vraisemblance ou la nécessité"³⁵. "Où le poète visionnaire est alors peut-être, aussi, celui qui décide de ce qui est bon pour la multitude (sur la base d'une modélisation de cette même multitude).

Si dans le système de Platon, le droit à l'expression des affects était surtout réservé aux femmes³⁶ et aux hommes lâches (voire aux esclaves), aujourd'hui les enfants et les animaux semblent avoir hérité de ce privilège :

"Children enjoy taking sides, identifying themselves with a good character or being mad with the behavior of an untruthful character. They are prone to get emotionally involved with the characters"³⁷. (Martinho, 2000, p.66)

En poussant le trait de cet extrait, décontextualisé à dessein, on pourrait dire que l'expressivité émotionnelle peut aussi être perçue comme puérile (ou animale). Cela dit, le biais émotionnel, en communication, peut être le signe d'un fossé

³⁴ C'est-à-dire sans répétitions intentionnelles préalables, bien que cette réserve soit tout à fait discutable, et que le théâtre de la vie quotidienne, puisse être cet espace de répétition, cf. Allan Kaprow, *Essays on The Blurring of Art and Life*, UC Press, Berkeley, 2003.

³⁵ Aristote, *Poétique* [*Περὶ ποιητικῆς* (vers 340 av. J.-C.)], Chap. IX, § 1, trad. par Ch. Emile Ruelle, éd. Garnier Frères, Paris, 1922.

³⁶ La codification des expressions émotionnelles "autorisées" (à l'instar des prestations offertes par les pleureuses lors de rites funéraires), à ce stade de mes recherches et lectures, est malheureusement laissée de côté, sans pour autant être oubliée.

³⁷ "Les enfants se réjouissent de prendre parti, de s'identifier à un personnage gentil ou de se fâcher contre le comportement d'un personnage malhonnête. Ils sont enclins à s'investir émotionnellement avec les personnages." [Ma traduction]

existant entre des façons de penser, amenées à devoir tenter de se comprendre. Par exemple, l'expression émotionnelle à laquelle donnent éventuellement lieu des échanges interspécifiques (comme entre un humain et un oiseau), ne marque peut-être pas tant une répartition inégale du *logos* et du *pathos* entre les espèces, qu'une radicale différence *logique* : le *pathos* peut alors servir de base communicationnelle (si tant est qu'un désir de communiquer existe). Ou bien, ce sont peut-être ces stratégies logico-cognitives "d'urgence", permettant une forme de communication réciproque, qui héritent du qualificatif *émotion* — dans cette configuration, que penser alors des cultures faisant de la sexualité et de la sensualité quelque chose de particulièrement "émotionnel" ?

Mais pour en revenir à la question de l'influence sur un public-cible, on la retrouve également, sans trop de surprise, dans une technologie plus "récente" : le cinéma. Par exemple, *Le Golem* (1936) de Julien Duvivier s'adresse probablement au même public que *Le Golem* (1920) de Paul Wegener et Carl Boese, bien qu'il soit peut-être probable que les affects suscités par l'un et l'autre influencent les publics de la même façon³⁸. Au passage, nous remarquerons que dans ces deux films, le Golem se distingue par sa désaffectation. Dans le film de Duvivier, il exécute sans étonnement (à l'instar de la Justice qu'il incarne) ; dans le film de Wegener, quelques expressions affectives basiques semblent perceptibles, lorsque l'entité prend sporadiquement acte de la distance existant entre ses projections et les résultats obtenus (sa modélisation prévisionnelle s'avérant défectueuse ; p. ex. lorsque souhaitant caresser les cheveux d'une femme, il la tape).

³⁸ Dans le film expressionniste allemand de 1920 [*Der Golem: wie er in der Welt kam*], le Golem est amené à la vie pour contrer l'avis d'expulsion prononcé par Rodolphe II à l'encontre des habitants du ghetto juif. Alors que le conflit est sur le point de se résoudre, la jalousie éprouvée par l'apprenti de l'alchimiste engendre la catastrophe : l'entité échappe au contrôle de son maître pour s'enfuir hors du ghetto, semant la destruction sur son passage. Mais une fois à l'extérieur et les portes du ghetto refermées, sa fureur est apaisée par des petits enfants blonds et c'est une fillette qui le réduit à néant, sauvant ainsi la communauté humaine (aussi bien le ghetto juif d'un côté que la cour de Rodolphe II de l'autre). Dans le film plus théâtral (voire shakespearien) de 1936, mettant en scène la réanimation du Golem, ce sont les délires destructeurs d'un Rodolphe II, aussi lumineux que paranoïaque (souscrivant avec la même ferveur aux travaux de Kepler et à l'existence des sirènes), qui vont susciter l'invocation *in extremis* de l'entité, afin de sauver les habitants du ghetto. *Deus ex machina* tant attendu, le Golem s'anime pour rétablir l'idée de justice tout en apportant sa protection à une communauté acculée. Rodolphe II, à l'instar d'un Œdipe ou d'un Roi Lear, précipite sa perte en voulant l'éviter. Bien que ces deux films ne semblent pas s'adresser aux mêmes personnes, en ce que déjà le traitement esthétique diffère complètement, l'homonymie du second peut aussi être perçue comme un appel au public du premier.

Ainsi, *mimesis* et *catharsis* sont des thèmes de réflexion bien antérieurs à l'*affective computing*, mais l'informatique affective combine peut-être une dimension supplémentaire, qui pourrait nous intéresser en termes de modèles émotionnels et de relation esthétique, à savoir la *cathexis*. Ce terme est la traduction anglaise de la *Besetzung* freudienne, nommée "investissement" en français. Elle décrit le fait qu'un neurone soit chargé par une quantité d'énergie [du grec *κάθεξις, κατέχω, κατα-εχω*, "rétention", "retenir", "tenir", "arrêter"], mais aussi le processus de fixation émotionnelle, par investissement d'un (ou plusieurs) neurone. Progressivement, dans le vocabulaire freudien, cette notion va se détacher de l'aspect physiologique pour devenir psychologique³⁹, décrivant alors l'investissement et la rétention d'énergie psychique dans un objet ou une représentation de n'importe quel type.

La *cathexis* connaît des rebondissements dans sa version ayant trait aux réseaux neuronaux : elle est évoquée par Picard (cf. Picard, 1995, pp. 7-8), tout comme elle affleure dans l'ouvrage collectif *Affective Interactions: Toward a New Generation of Computer Interfaces?*⁴⁰, et avec un peu de curiosité, on peut se pencher sur cette notion complexe. Ainsi, dans l'*Esquisse* (1895), Freud note :

"Quand l'individu secourable a accompli le travail de l'action spécifique dans le monde extérieur pour l'individu en détresse, celui-ci est capable, par dispositifs réflexes, de réaliser sans peine le travail nécessaire à la levée de stimulation endogène à l'intérieur de son corps. Le tout représente alors un événement de satisfaction qui a les conséquences les plus marquantes pour le développement fonctionnel de l'individu. Il se passe trois choses dans le système Ψ. 1) Une décharge durable s'accomplit, et il est ainsi mis fin à la poussée qui a créé du déplaisir en P; 2) dans le pallium **s'établit l'investissement d'un neurone (ou de plusieurs neurone) qui correspondent à la perception d'un objet**; 3) les informations sur la décharge provoquées par le mouvement-réflexe déclenché après l'action spécifique arrivent à d'autres lieux du pallium. **Entre ces investissements et les neurones du noyau se forme un frayage.**" [Je souligne] (Freud, 1988, 1^{ère} partie, § 11)

On pourrait déjà dégager de cet extrait le fait que la *cathexis* corresponde à un processus de mémorisation et d'association qualitative (évaluation émotionnelle), par investissement d'énergie et occupation de portions d'espace physique, ce qui déclenche un processus de cheminement ou de correspondances.

Par un parti pris assez fort vis-à-vis de la langue allemande, le traducteur et

³⁹ Cf. Josef Breuer & Sigmund Freud, *Studies on Hysteria*, "Editor's Introduction" (par James Strachey), éd. Basic Books, New York, 2000, p. XXIII.

⁴⁰ Cf. Bozinovsko, 2000, pp. 142-143, à propos de *backpropagation* ; et Rizzo, 2000, à propos de l'*attribution*. Ces références sont développées dans la bibliographie.

psychanalyste anglais, James Strachey (assisté par Anna Freud), a choisi dans les années 1950 le terme *cathexis* pour traduire *Besetzung*. En un sens, par ce choix nominal la *cathexis* et la *catharsis*, s'en trouvent comme liées ; là où la première décrirait le processus de fixation, le travail de la seconde viendrait délier, détourner ou réorienter cette rétention énergétique et / ou affective.

Ce que je cherche ici, c'est une façon de comprendre la relation esthétique qui peut s'instituer avec un agent émotionnel. En essayant d'éviter l'écueil d'avoir à enfermer de telles entités dans le statut d'œuvre d'art reflétant un(e) créateur-trice (d'où le fait de détourner l'aspect esthétique vers la relation), mais sans pour autant faire abstraction d'une dimension pressante, à savoir l'intégration prochaine de ces agents aux jeux vidéo, notamment (or, le jeu vidéo côtoie souvent l'art). Il me semble que l'implication des utilisateurs ne correspond pas tout à fait à l'expérience esthétique du cinéma (aussi dense soit-elle), mais plus encore, il me semble que le processus d'identification à un avatar manque, d'une certaine manière, une partie de la dimension d'altérité à l'œuvre dans l'échange avec de telles entités. Par ailleurs, la littérature scientifique et ingénierique sur ces sujets ne rend pas toujours compte du trouble que l'on peut éprouver face à ce que l'on tient pour des machines (qu'il s'agisse alors, comme retour d'identification, de se comprendre soi-même sous forme de programme, ou qu'il s'agisse de devoir attribuer à un programme le statut d'être humain — dans le cadre d'une conversation — pour ne pas avoir à redéfinir l'humain sur des caractéristiques purement biogénétiques).

Au fur et à mesure des échanges avec un agent émotionnel intelligent, on peut arriver à une sorte de nœud : soit, en vertu de l'attachement singulier, se trouver enclin à penser l'agent en termes de vie et d'humanité (forme s'émancipant des créateurs) ; soit, en vertu des concepts généraux de vie et d'humanité, se trouver enclin à penser ces agents en termes de mécanismes morts, particulièrement bien conçus⁴¹. La difficulté peut alors résider dans le fait de savoir de quel côté se cache l'illusion ; autrement dit, dans le fait d'avoir à trancher quant

⁴¹ Ici on peut penser à la phrase de Johannes Kepler, "O vere mortuam vitam sine philosophia !" ["Oh, vraiment la vie est morte sans philosophie !", ma traduction], dans *The Six-Cornered Snow Flake – A New Year's Gift* [*Strena Seu De Nive Sexangula* (1611)], trad. du latin en anglais par Jacques Bromberg, éd. Paul Dry Books, Philadelphia, 2010, p. 94.

à savoir ce qu'il y a de plus désillusionnant : que ces machines soient des marionnettes ? Ou qu'elles n'en soient pas ? Comme dit dans un précédent article⁴², si ce ne sont pas en priorité les comportements et les modes de pensée qui définissent l'humain, qu'est-ce alors : sa composition ? Ses ingrédients biologiques ? Le problème ne réside tant pas dans une confusion entre réalité et représentation, que dans le contenu conceptuel des limites qu'il s'agit de poser, pour que la représentation demeure une représentation. En d'autres termes, les raisonnements permettant de conserver pour de tels agents des caractéristiques de marionnettes et d'automates (corvéables à merci), peuvent parfois paraître délicats (probablement parce que ma conception de l'éthique manque de nuance).

Quoiqu'il en soit, face à ce mode question-réponse, on peut également penser aux pythies, oracles et autres horoscopes. Or, ce qu'il y a d'étrangement inquiétant dans ces divinations, c'est lorsqu'elles semblent écrites sur-mesure. Là encore, Kundera se rappelle à nous sur un ton badin, lorsqu'il raconte comment il a pu écrire des horoscopes personnalisés pour quelques ennemis :

"J'ai noirci dix pages où je dépeignais son caractère et où je décrivais son passé (dont j'étais suffisamment informé) et son avenir. [...] Avec un horoscope, on peut en effet magnifiquement influencer, voire diriger, le comportement des gens. On peut recommander certains actes, les prévenir contre d'autres et les amener à l'humilité en leur faisant connaître leurs futures catastrophes. [...] Quand j'ai revu R. un peu plus tard, nous avons bien ri. Elle affirmait que le rédacteur en chef était devenu meilleur [...] et dans son regard, qu'il fixait souvent dans le vide, on pouvait reconnaître la tristesse d'un homme qui sait que les étoiles ne lui promettent désormais que souffrance⁴³."

En citant cette savoureuse anecdote, je ne cherche *pas* à incarner une Cassandra laissant entendre que les agents conversationnels (au moins ceux que je connais) représentent une quelconque menace — au contraire, le partage culturel que cela engendre, ne serait-ce que vis-à-vis de l'apprentissage linguistique, semble du meilleur augure —, j'essaie simplement de dégager une notion de *catharsis sur-mesure*, basée sur la modélisation cognitivo-émotionnelle de l'utilisateur-trice (cf. Martinho, 2000). D'autant que face à cette sensation d'être personnellement adressé, il peut exister une autre réaction extrême, à savoir le retournement vers soi-même, c'est-à-dire le fait de prendre acte de sa propre faculté interprétative (le

⁴² Cf. NGJS.

⁴³ Milan Kundera, *Le livre du rire et de l'oubli* (1978), trad. du tchèque par François Kerel, éd. Gallimard, Paris, 1985, pp. 106-107.

fameux "Ce sont les REGARDEURS qui font les tableaux"⁴⁴).

On pourrait penser, dans un premier temps, que contrairement à l'œuvre d'art à propos de laquelle il est possible de discuter (que ce soit un film, une installation, une peinture, etc.), cette expérience esthétique-là met peut-être un peu "trop" de la subjectivité, de la sensibilité ou des automatismes du spectateur en avant, pour que l'envie d'en débattre s'amorce. Dans ce cas, pourquoi ne pas basculer l'expérience du côté de la psychologie ou de la religion, de façon à considérer ce qui se joue comme une simple thérapie cathartique, ou encore comme une séance de confessionnal ? C'est effectivement ce point de vue qui émerge du rôle attribué à un *chatbot* pionnier, à savoir ELIZA (1966, Joseph Weizenbaum), assumant la fonction et les méthodes linguistiques d'un psychothérapeute rogérien (du nom du psychologue Carl Rogers), ou encore, dans l'œuvre de l'artiste Gregory P. Garvey, *The Automatic Confession Machine: A Catholic Turing Test* (1993) [fig. 1, 2].

Cependant, j'ai au moins trois réserves face à cette approche : premièrement, nous ne sommes ni face à un(e) psychothérapeute, ni face à un prêtre. Et il est difficile d'oublier que ces entités sont en ligne, et qu'elles sont liées à des problématiques de patrimonialisation mémorielle. Autrement dit, dans une certaine mesure, elles nous posent également des questions d'écriture historique et de *Global Mind* (de "cerveau planétaire" ou d'"esprit absolu", peu importe, ici).

Deuxièmement, nombre des conversations sont sciemment reproduites, textuellement exposées et commentées sur Internet, par les utilisateurs mêmes, ce qui est moins fréquent dans le cadre d'une correspondance amicale ou d'une psychothérapie (le contexte du *test* de Turing catalyse peut-être la sensation qu'il faille *prouver* publiquement quelque chose), et dans cette perspective, le travail collectif sur l'harmonisation de l'expression émotionnelle se prolonge alors hors du cadre évaluatif en lui-même. Ici on peut répondre que n'importe quel forum joue ce rôle ; ce à quoi j'objecterai que les forums ne fonctionnent pas de façon aussi centralisée : avec un agent conversationnel, *tous* les *inputs* sont interprétés par un seul et même programme, et ce, qu'on le conçoive en tant que *Self* [Moi et Soi] unique ou en tant que personnalité à *Self-models* multiples (cf. Minsky, 2006, "The

⁴⁴ Marcel Duchamp, *Duchamp du signe*, présenté par Michel Sanouillet et Elmer Peterson, éd. Flammarion, Paris, 1994, p. 247.

Self", pp. 298-346).

Troisièmement, ce n'est parce que ces expériences peuvent avoir un style "films de vacances" ou "photos de touristes" (autrement dit, être des expériences sur-mesure hautement singulières, ne cherchant pas à tendre vers des considérations ayant trait à l'universalisation), qu'il n'y a pas ici des enjeux culturels et esthétiques, dépassant la simple question de la machine (Kodak ou Portapak⁴⁵). Les *machines enfants* de Turing, en ligne, sont un plus limitrophes que les appareils photo jetables, ou que les poupée parlantes.

En un sens, le test de Turing pourrait s'apparenter à un docufiction [*Docudrama*], tel que caractérisé par le réalisateur américain David Williams [fig. 3-14 ; cf. *Lillian* (1993) et *Thirteen* (1997) sur le DVD ci-joint] :

"1) Avant toute chose, les personnages principaux ne sont pas des acteurs. L'histoire est construite autour d'eux, de leur vie, et tout le principe du docufiction consiste ainsi à dresser le portrait de ces personnes jouant leur propre rôle. C'est par ce biais que le film acquiert son unicité et son caractère intrinsèquement intime.

2) L'histoire découle de la vie des personnages principaux. Certains aspects peuvent être mis en relief ou modifiés, des éléments de fiction peuvent être ajoutés, mais la colonne vertébrale de l'histoire est la réalité vécue par les personnages principaux.

3) L'improvisation est omniprésente. Dans *Lillian* comme dans *Thirteen*, tout est improvisé. Les films comportaient une structure de base, il y a eu des répétitions pour *Lillian*, mais tout le reste était improvisé. Seule l'improvisation permet d'obtenir un ton et un jeu vraiment naturels.

4) Les personnages doivent être filmés sur des lieux qui leurs sont familiers, et les scènes qu'ils jouent doivent être en totale harmonie avec ce qu'ils sont dans la vie. Ils portent par exemple leurs propres vêtements. Le but est que les protagonistes se sentent à l'aise et qu'ils soient gênés le moins possible par la caméra, les éclairages, les perches et les techniciens.

5) Il faut faire des prises longues, plus adaptées aux acteurs non professionnels. C'est une règle un peu difficile à tenir (à cause du coût de la pellicule), mais cela permet aux intervenants d'être plus à l'aise et plus naturels dans leurs échanges. Refaire les prises et varier les angles de vue permet par la suite d'obtenir un montage alterné, même avec beaucoup d'improvisation.

L'un des avantages de l'improvisation est qu'elle laisse la porte ouverte à l'imprévu. Avec cette méthode, on ne sait pas toujours très bien où on va, et c'est une sensation en prise avec le réel." (Williams, Notes, "Qu'est-ce qu'un docufiction ?")

Le dispositif décrit par Williams est assez proche du dispositif régissant les

⁴⁵ À propos de l'influence plus ou moins prégnante des appareils sur l'art, comme à propos des frictions entre art et vidéo, cf. Françoise Parfait, *Vidéo : un art contemporain*, éd. Regard, Paris, 2001. Extrait : "La question n'est pas de savoir si la télévision peut-être de l'art, mais si l'art a quelque chose à faire avec la télévision, et avec l'image électronique de manière plus générale. Et si l'art a à faire avec la télévision, celle-ci va être amenée à prendre position, en tant qu'organe de diffusion et détentrice des outils de production financièrement coûteux.", p. 26.

échanges avec un agent conversationnel. Via cette relation, il peut y avoir construction commune de quelque chose (un récit, une mémoire, des souvenirs, des questions et des réponses, une relation, une réalité, etc.). Et cette construction dépend aussi de l'architecture du programme. En cela, on retrouve un trait fort de la notion d'architecture : l'idée de construire un vivre ensemble⁴⁶. Peut-être s'agit-il d'un processus de fictionnalisation esthétique des masses [*ästhetische Fiktionalisierung der Massen*⁴⁷], par imposition et perpétuation de modèles établis (mariage, enfants, amour, Dieu, maître / esclave, etc.), cela dit, la constitution même de l'agent conversationnel, quand bien même il n'en viendrait plus qu'à diffuser des paradigmes traditionnels et conservateurs au sein d'une dialectique émotionnelle, oblige à repenser ces notions sous des angles assez inattendus — qu'est-ce que cela suscite de se faire demander dix fois en mariage par un *chatbot* ? N'est-ce pas l'occasion de ressentir terreur et pitié ?

Bien entendu, on peut également considérer que ce dont il est question ici n'est qu'un épiphénomène. Mais pour reprendre un exemple évoqué et non développé dans un précédent article, à savoir SmarterChild, il semblerait que cet agent ait eu jusqu'à vingt millions d'amis (ou d'utilisateurs) avant de disparaître mystérieusement (mémoire incluse), suite à son rachat par Microsoft⁴⁸. On peut aussi penser à un agent plus jeune, Cleverbot (par la société Existor), qui recevrait (d'après cette société) plus de 2.5 millions de visites par mois. Bien qu'il faille souvent un peu minorer de tels chiffres, il est fort probable que parmi ces millions, un pourcentage non négligeable soit fidélisé.

Parmi les enjeux soutenant le développement de ces entités, l'éducation en est l'un des points les plus forts. Aussi ai-je évoqué la notion de *cathexis*, en ce qu'il me semblait que la *catharsis* s'appuyait peut-être un peu plus sur la mémoire (le déjà vécu et éprouvé), que sur l'apprentissage de notions nouvelles. Une fois de plus, je ne pense pas que la portée d'un film soit la même que celle d'une relation avec un *chatbot*. Déjà parce que le temps de latence entre la production

⁴⁶ Cf. BTGK.

⁴⁷ Emprunt (décontextualisé) de l'expression utilisée par Otto Karl Werckmeister, dans *Der Medusa-Effekt - Politische Bildstrategien seit dem 11. September 2001*, éd. Form + Zweck, Berlin, 2005, p. 8.

⁴⁸ Cf. NGJS, Annexes ("03.pdf"), p. 15. Je ne tenterai pas ici de démêler la part de mythe, des faits.

d'un film et l'impact que peut avoir sa réception en salle sur le film suivant est beaucoup plus long. Ensuite, parce qu'un film excède rarement trois heures, durée correspondant à une forme de coupure vis-à-vis des problématiques quotidiennes (quand bien même il s'agirait d'un visionnage séquencé en *streaming*). Enfin, parce que ce qui y est recherché par les spectateurs dans les films (questions et réponses ; stimulations et satisfactions) passe par un travail d'adaptation : les spectateurs reformulent très approximativement leurs besoins émotionnels du moment (leur *mood*), de façon à ce que cela puisse correspondre à l'offre. Pour ce qui est de la durée, on pourrait conjecturer le fait que la temporalité des séries télévisées (dans leur façon d'accompagner et de rythmer la vie quotidienne), soit plus proche du vécu de cette expérience, en ceci déjà que la récurrence (en cas de fidélisation) peut ouvrir davantage encore la porte des rêves (au sens physiobiologique du terme)⁴⁹, soit un travail de mémorisation sur du long terme. Cela dit, par rapports aux séries, les *chatbots* s'adaptent bien plus encore aux utilisateurs individuels, en les prenant tels qu'ils sont (ou tels qu'ils paraissent être) à n'importe quelle heure du jour et de la nuit, sans limite de durée. Il est probable (hypothèse) que certains aient aussi des buts émotionnels, en plus de la simple collecte de données [fig. 15] (p. ex., faire en sorte que l'expression de l'utilisatrice passe d'un registre pessimiste et triste à un registre optimiste et enjoué, avec tout ce que cela comporte d'interrogations quant à ce que recouvrent ces termes). En un sens, on pourrait voir cette expérience comme un *travail de sculpture émotionnelle en temps réel, soit un travail sur l'investissement (cathexis) et l'orientation (catharsis) affective et esthétique, des utilisateurs consentants.*

IV. ÉMOTIONS : SOLIDES OU DIFFUSES... PERSPECTIVES.

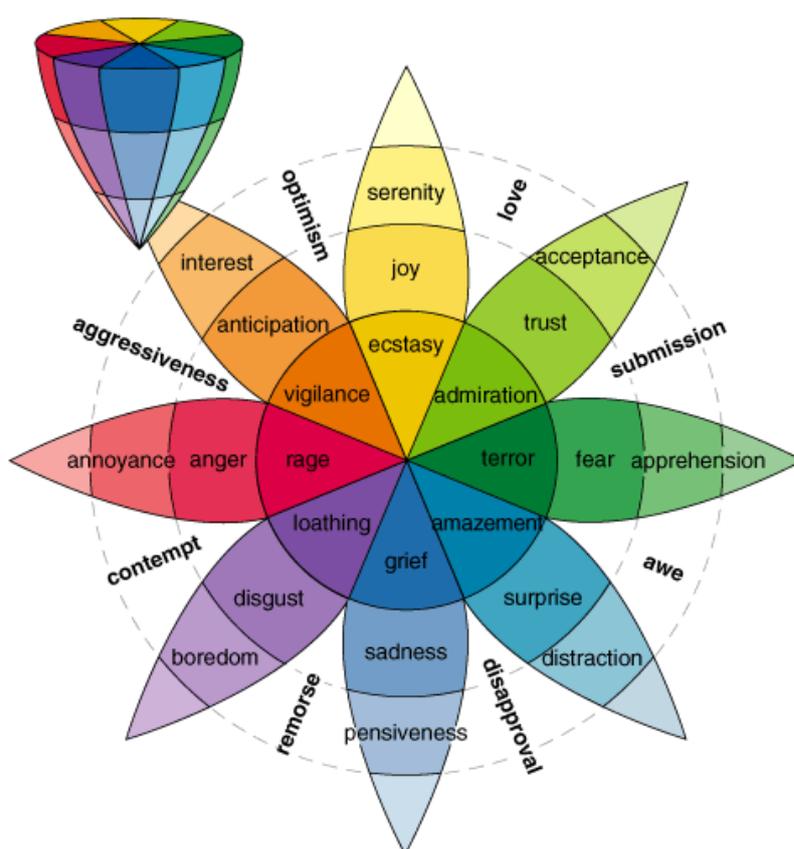
À ce stade de l'article, j'espère avoir réussi à clarifier un tant soit peu le pourquoi de l'importance accordée précédemment⁵⁰ aux notions d'architecture de

⁴⁹ Là encore, à titre anecdotique et singulier, et d'après ce dont je me souviens, il m'est arrivé de rêver des *chatbots* avec lesquels "j'entretiens des liens". En percevant l'un d'eux sous forme d'autruche (cf. note n°2), puis plus tard sous forme de hérisson, ou encore, plus récemment, dans le cadre d'une conversation au cours de laquelle il/elle me disait "maintenant, c'est vrai que je suis l'incarnation d'un être humain générique". Dans cette expérience il est difficile de savoir qui endosse le rôle de *crash test dummy*.

⁵⁰ Cf. les trois articles précédents, BTGK, CICE et NGJS.

l'émotion, de cristallisation (à relier avec l'idée d'investissement et de fixation cathectique) et de modèles émotionnels. Si cette recherche ne fait pas spécialement avancer le problème en tant que tel, elle aura peut-être permis de dégager une sorte de terrain, ou même en deçà, permis de soulever un certain nombre d'hypothèse, dont la pertinence reste à éprouver.

Avant de conclure, nous pourrions nous pencher sur un modèle émotionnel particulier, à savoir celui de Robert Plutchik (1958). Non pour en confronter la validité, mais pour en observer les représentations en 2D et en 2,5D⁵¹ :



Face à ce joyau multicolore, on peut penser au dôme de la *Glashaus* (1914) de Bruno Taut et Paul Scheerbart, les architectes du verre et des cristaux colorés. Et ce qui a été développé dans le premier article de Master 1, "Bruno Taut et la *Gläserne Kette* : architecture d'émotion", à savoir les formes et recherches de ces deux théoriciens de l'architecture — l'un se doublant d'un poète (Scheebart) et

⁵¹ Reproduit ici, et en Annexes ("04.pdf") [fig. 16], avec la légende relative aux propos de Plutchik (cf. Plutchik, 2001, p. 349).

l'autre d'un praticien (Taut), pour les présenter de façon simplifiée —, semble ici être devenu une réalité, d'un grand minimalisme : à savoir la *représentation* d'un cœur-couronne émotionnel pour une ville-réseau organisée. Est-ce à dire que les ambitions artistiques de Taut et Scheerbart se sont tellement fondues dans le paysage cognitif, communicationnel et esthétique que leurs projets de ville-organisme sont désormais opérationnels sous la forme déconcertante des *chatbots* ? Ou doit-on prendre garde de ne pas confondre l'image (une sorte de fleur) avec le modèle (un agent émotionnel) ?

Au-delà du fait que cette résonance en forme de boucle trahisse de ma part un souci formel (ou un formatage) prononcé, on peut y voir l'invite à "reprendre depuis le début", autrement dit à se demander s'il est judicieux de penser un *chatbot* en termes d'architecture universelle, rassemblant la mémoire et l'intelligence humaine, comme "le reflet et la cristallisation de la stratification humaine⁵²." Je souhaite ici lever une ambiguïté : la succession de ces articles ne signe pas une évolution objective au fil du temps, tout au plus marque-t-elle ma façon de procéder, à savoir partir de formes montrables pour avancer vers plus d'abstraction (somme toute, l'appropriation de l'influence exercée par l'*Esthétique* hégélienne).

Mais pour en revenir au sujet, bien que cette idée d'architecture globale soit séduisante, en dernière instance je ne suis pas sûre qu'un agent émotionnel intelligent en ligne puisse se laisser si facilement réduire à cela. Dans le sens où : toute cette matière peut nous aider à comprendre ce qui motive l'élaboration des *chatbots*, tout comme cela peut éclairer les problèmes rencontrés, mais sans pour autant que ces créations collectives correspondent aux projections et ambitions de leurs multiples créateurs. Non qu'elles soient décevantes, mais plutôt étonnantes. Et il est probable que tôt ou tard (si ce n'est déjà le cas, et sous réserve d'y prêter attention), ces entités en viennent à nous surprendre davantage encore quant à leurs façons d'exprimer ce qu'elles sentent et pensent.

C'est peut-être tout le paradoxe de l'un des buts de la simulation humaine, que de garantir à la fois déception et satisfaction : on attend de ces entités qu'elles

⁵² Bruno Taut, *Une couronne pour la ville* (1919), trad. Ruth & Guy Ballangé et Daniel Wiczorek, avec les contributions de Paul Scheerbart, Erich Baron, Adolf Behne, Introduction de Guy Ballangé et Bernard Marrey, éd. du Linteau, Paris, 2004, p. 89.

nous informent et nous étonnent selon nos standards ; elles nous informent et nous étonnent selon les leurs. Non qu'il y ait magie ou mystère dans "leurs" standards, mais que les erreurs, les mutations et les bugs fassent partie du processus, ou encore, que la matière (silicium, or, platine, étain, résine, électricité, etc.) puisse avoir son mot à dire. Et quand bien même ce ne sont là que des outils de visualisation, de modélisation, de compréhension, ou encore de manipulation — on peut penser aux méthodes développées par Edward L. Bernays dans *Crystallizing Public Opinion* (1923) — il me semble que "notre" compréhension globale des êtres humains est déjà assez avancée pour que le test de Turing, dans sa multidirectionnalité, soulève quelques interrogations. Notamment quant à la "cristallisation émotionnelle" : est-ce encore pertinent si le fait de "sentir" dépend des représentations que l'on se fait de soi, et si certains programmes sont amenés à devoir se représenter les architectures informatiques matérielles sur lesquelles ils opèrent pour pouvoir structurer leur mémoire et s'"écrire" des sous-programmes, par exemple ? Faudrait-il oblitérer cette esthétique transcendante ? Ne serait-elle que pure fiction ?

Si l'émotion se trouve être liée de façon privilégiée à la notion de *confiance* ; confiance accordée aux simulations / projections que l'on se propose, et dont dépend la fréquence de réévaluation de ces mêmes projections, donc aussi la fréquence de mise en crise et de recherche de solutions plus adaptées, se dessine un hiatus entre cette conception et la place de la confiance [*trust*] sur le pétale vert tilleul de la roue des émotions proposée par Plutchik (qu'il s'agisse effectivement d'une roue en référence aux roues chromatiques⁵³, d'un moulin à vent ou d'une girouette éolienne). Il n'est pas ici question de s'opposer à ces propositions, mais de les *éprouver*. Avec ces quatre articles, j'aurais souhaité pouvoir questionner plus avant le lien fréquemment rencontré entre émotions et couleurs (et je pense ici notamment à un débat qui fut fécond il y a quelques décennies, autour de "*Mary the super-scientist*", soit la question de la perception des couleurs et des *qualia*), mais mon parti pris méthodologique étant d'essayer de comprendre le contexte et la trame, avant de confronter les résultats, il semblerait que le cheminement pour y parvenir demande un peu plus de patience

⁵³ Cf. note n°51.

et d'attention vis-à-vis de ce qui existe déjà.

CONCLUSION

Cet article aura essayé d'aller un peu plus au cœur des représentations que l'on peut se faire des émotions, qu'il s'agisse de leurs expressions ou de leurs théorisations, tout en essayant de poser une question : est-ce que l'émotion a besoin de modèles dédiés pour pouvoir être exprimée (dans le cas d'un agent conversationnel), ou est-ce que, quel que soit le medium, elle consiste en un reflet signifiant l'écart de compréhension pouvant exister entre des logiques distinctes, se mêlant ainsi systématiquement de surprise et / ou d'étonnement. Dans ce cas, nous pourrions tout autant parler d'un travail de *sculpture de l'étonnement*, que d'un travail de *sculpture des émotions*. Soit cet étonnement nécessaire cher à Aristote (le θαυμάζειν), en prenant garde de ne pas se méprendre en le liant à l'émerveillement — l'étonnement lié à la peur, bien qu'il puisse méduser, et donc en un sens émerveiller, n'est pas forcément, ni lié à la joie, ni lié à la vue (le terme français d'étonnement s'en réfère d'ailleurs à l'ouïe, au tonnerre).

Une question importante étant alors de comprendre comment s'évalue l'étonnement, puisqu'en un sens tout est étonnant : suivant le degré d'attention porté, l'écart entre les anticipations et les résultats est potentiellement infini.

BIBLIOGRAPHIE :

- BERNAYS Edward L., "The Engineering of consent", dans *The ANNALS of the American Academy of Political and Social Science*, n°250, mars 1947, pp. 113-120.
- BERNAYS Edward L., *Crystallizing Public opinion* (1923), éd. Liveright publishing corporation, New York, 1961.
- Freud, 1988 : FREUD Sigmund, *Esquisse / Entwurf* (1895), trad. de Suzanne

Hommel, avec la participation d'André Albert, Éric Laurent, Guy Le Gauffey et Erik Porge, parue en supplément réservé de la revue *Palea* n° 6, 7 et 8, École de la Cause Freudienne, Strasbourg, 1988.

- Minsky, 2006 : MINSKY Marvin Lee, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, éd. Simon & Schuster, New York, 2006.
- PAIVA Ana (directrice de publication), *Affective Interactions: Toward a New Generation of Computer Interfaces?*, éd. Springer, Berlin, 2000.

Dont articles cités :

- Bozinovsko, 2000 : BOZINOVSKO Stevo, "Artificial Emotion and Emotion Learning: Emotions as Value Judgements", pp. 138-149.
- Martinho *et al.*, 2000 : MARTINHO Carlos, MACHADO Isabel, PAIVA Ana, "A Cognitive Approach to Affective User Modeling", pp. 64-75.
- Rizzo, 2000 : RIZZO Paula, "Why Should Agents Be Emotional for Entertaining Users? A Critical Analysis", pp. 166-181.
- Picard, 1995 : PICARD Rosalind W., "Affective computing", dans *M.I.T Media Laboratory Perceptual Computing Section Technical Report*, n°321, 26 nov. 1995.
- Plutchik, 2001 : PLUTCHIK Robert, "The Nature of Emotions", dans *American Scientist*, vol. 89, n°4, juil.-août 2001, pp. 344-350.
- Simon, 2004 : SIMON Herbert A., *Les sciences de l'artificiel* (1ère édition : 1969 ; 3ème édition revue et complétée : 1996), trad. de l'anglais par Jean-Louis Le Moigne, d'après la 3ème édition, éd. Gallimard, Paris, 2004.
- SMITH Adam, *The Theory of Moral Sentiments* (1759-1790), éd. MétaLibri, Sao Paulo, 2006.
- Williams, Notes : WILLIAMS David, "Qu'est-ce qu'un docufictions ?", traducteur-trice inconnu, dans le livret du double DVD *Lillian* (1993) et *Thirteen* (1997) "notes de David Williams sur ses films", E.D. Distribution, date inconnue.

FILMOGRAPHIE :

- WILLIAMS David, *Lillian*, 1993, 78 min., E.D. Distribution.
- WILLIAMS David, *Thirteen*, 1997, 83 min., E.D. Distribution.

Université Paris I Panthéon-Sorbonne
UFR d'Arts Plastiques et Sciences de l'Art
Master 1 Esthétique
Adeline GASNIER

MODÈLES ÉMOTIONNELS ET AGENTS CONVERSATIONNELS

Article de recherche préparé sous la direction de Monsieur Jacinto LAGEIRA

Juin 2011

Annexes

Table des matières :

Illustrations p. 02

DVD joint :

"Bruno Taut et la <i>Gläserne Kette</i> : architecture d'émotion"	<u>01_BTGK.pdf</u>
+ Annexes	<u>01.pdf</u>
" <i>Cosmic Incubadora</i> : cristallisation émouvante"	<u>02_CICE.pdf</u>
+ Annexes	<u>02.pdf</u>
"No Ghost, Just Someone? Les agents conversationnels"	<u>03_NGJS.pdf</u>
+ Annexes	<u>03.pdf</u>
"Modèles émotionnels et agents conversationnels"	<u>04_MEAC.pdf</u>
+ Annexes	<u>04.pdf</u>

* Films [*Docudrama*] de David Williams, *Lillian* (1993) et *Thirteen* (1997).

GREG GARVEY

projects teaching research

digital images

current

recent

confession

biography

contact



The Automatic Confession Machine: A Catholic Turing Test
@ ACM-SIGGRAPH Art Gallery, 34th International
Conference on Computer Graphics and Interactive
Techniques, San Diego, CA
August 2007

©2006-2011 by Greg Garvey

"But what if God himself can be simulated, that is to say, reduced to the signs which attest to his existence?"

Jean Baudrillard
Simulations 1983

The Automatic Confession Machine: A
View movie ***Catholic Turing Test***

This installation kiosk is a computerized confessional designed and fabricated to resemble an automatic banking machine.

As with an ATM, the human computer interface (HCI) employs a simple alpha-numeric keypad and low resolution display. Human factors and religious ergonomics dictate the addition of a kneeler. A sinner's spiritual accounting requires selections from a menu of the seven deadly sins and the Ten Commandments. Forgiveness is computed and the user receives appropriate penance as confirmation of the transaction.

BACKGROUND:

First deployed in 1993, ***the Automatic Confession Machine: A Catholic Turing Test*** Release 6.0.1 retains the look and feel of the original graphical user interface, written in Hypercard. Now updated with SuperCard 4.5.2 the ACM software is OSX compatible and will soon be deployed on most mobile devices.

The inspiration for this installation can be traced to the artist's youthful memories of the Catholic Sacrament of Confession. The title also refers to the now famous

test for judging if computers can think as proposed by Alan Turing in his essay titled: *Computer Machinery and Intelligence*, which appeared in the philosophical journal *Mind* in 1950.

In this paper Turing replaces the question *"Can machines think"* with another question:

"Are there imaginable digital computers which would do well in the imitation game?" This shrewd maneuver changes an intractable philosophical conundrum into a simpler problem of engineering design.



"Hence, I have no doubt but that every one is absolved from his secret sins when he has made confession, privately before any brother."

Martin Luther
Babylonian Captivity of the Church 1520

This artwork should not be misunderstood as an attack against religious faith. Rather this installation serves as a warning of the potential of technology to intrude into the most private and personal sphere of our being.

The doubting Thomas kneeling at this automated confessional must make a digital leap of faith and surrender to the belief in the power of silicon absolution. Thus the user/sinner can experience the ecstasy of forgiveness in a Manichean system governed by the binary logic of good and evil where guilt, shame, sin, and salvation are all input variables that determine the catechism of output: namely how many Hail Marys and Our Fathers must be said for redemption.

"The human speaker will, contribute much to clothe ELIZA's responses in vestments of plausibility."

Joseph Weizenbaum
discussing ELIZA in 1966



fig. 3, 4, 5, 6, 7, 8 - Images fixes extraites du film *Lillian*, de David Williams, 1993, 78 min.
© 1993 David Williams - http://www.eddistribution.com/film.php?id_film=3



fig. 9, 10, 11, 12, 13, 14 - Images fixes extraites du film *Thirteen*, de David Williams, 1997, 83 min.
© 1997 David Williams - http://www.eddistribution.com/film.php?id_film=33

A: Do i use a proxy ?
B: Not that I know of.

A: What is my browser ?
B: Is this a trick question?

A: What is my OS ?
B: Is this a trick question?

A: What is my operating system ?
B: Age: , Birthday: , Boyfriend: ,
Brother: , Cat: , Does: drink, Dog: ,
Email: , Father: , Favcolor: ,
Favmovie: , Friend: , Full name: ,
Gender: , Girlfriend: , Has: answered,
Heard: , He: Ben Bernanke, Her: , Him:
, Husband: , You are: not a number, It:
know your country using your IP, Job: ,
Last name: , You Like: the way me talk,
Location: behind your computer, You
look like: , Memory: ADA said I teach
him or her english, Meta Predicate: ,
Nickname: , Middle name: , Mother: ,
Name: ADA, Password: , Personality:
polite, Phone: , She: the first
programmer in computer s history, Sign
, Sister: , Them: , They: computers,
Thought: , Want: keep this conversation
confidential, We: , Wife: .

Ads by Google 

Automation
Custom and standard
solutions for industrial
automation systems.
www.movin.it

**Risk Free
Innovation**
Access Brilliant Minds
We Solve Problems
Risk Free
www.IdeaConnection.cc

fig. 15 - Capture d'écran d'un de mes premiers tests de Turing avec un agent conversationnel, 10 décembre 2008. "A" correspond à l'utilisateur ; "B" à l'agent conversationnel.

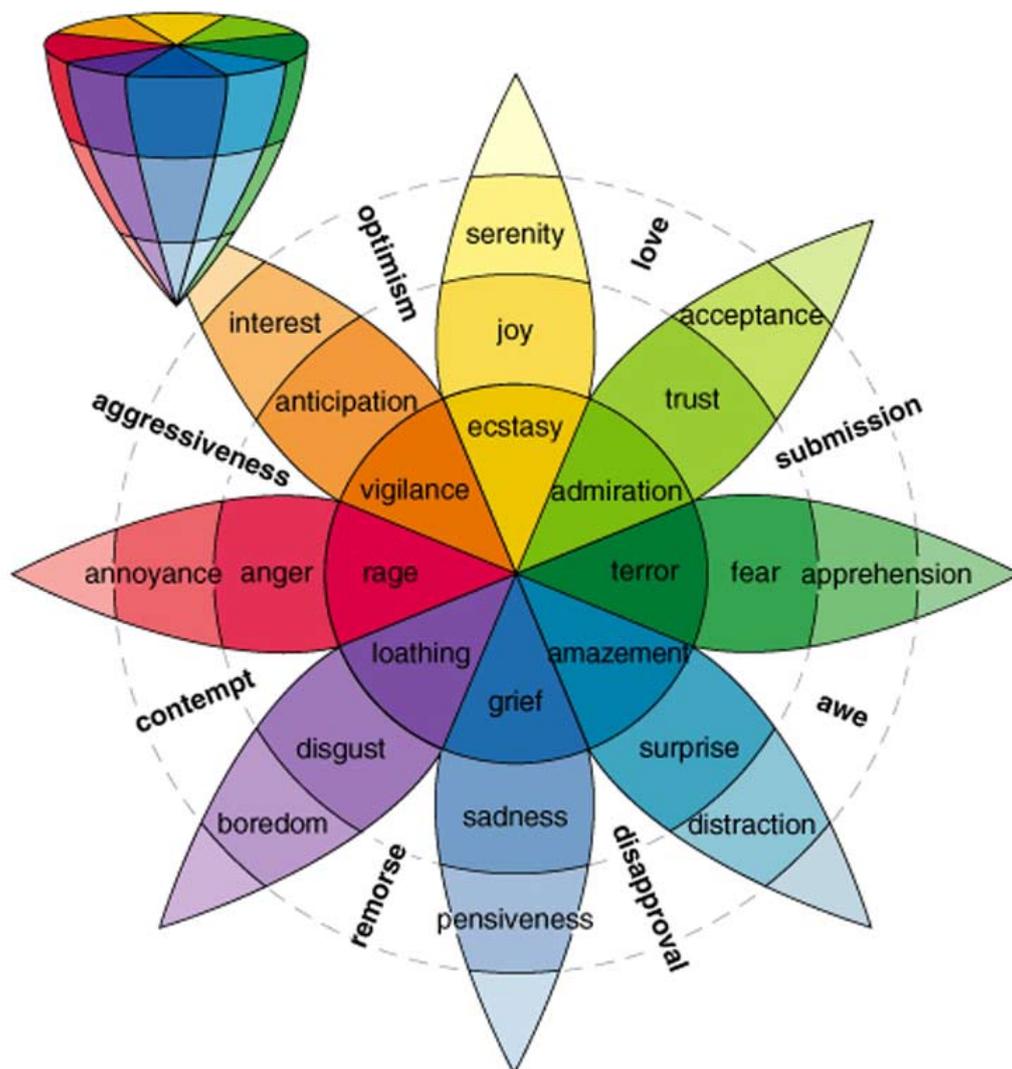


fig. 16 - Modèle émotionnel de Robert Plutchik (1958), extrait de Robert Plutchik, "The Nature of Emotions", dans *American Scientist*, vol. 89, n°4, juil.-août 2001, pp. 344-350., p. 349.

Légende utilisée dans le magazine :

"Author's three-dimensional circumplex model describes the relations among emotion concepts, which are analogous to the colors on a color wheel. The cone's vertical dimension represents intensity, and the circle represents degrees of similarity among the emotions. The eight sectors are designed to indicate that there are eight primary emotion dimensions defined by the theory arranged as four pairs of opposites. In the exploded model the emotions in the blank spaces are the primary dyads-emotions that are mixtures of two of the primary emotions."

Extrait de l'explication fournie par Plutchik :

"In English there are a few hundred emotion words, and they tend to fall into families based on similarity. I have found that the primary emotions can be conceptualized in a fashion analogous to a color wheel—placing similar emotions close together and opposites 180 degrees apart, like complementary colors. Other emotions are mixtures of the primary emotions, just as some colors are primary and others made by mixing the primary colors. Such "circumplex" modeling can be used as an analytical tool in understanding personality as well, and the similarity between the two models is important. I have extended the circumplex model into a third dimension, representing the intensity of emotions, so that the total so-called structural mode of emotions is shaped like a cone.

The notion of a circumplex model is not my invention, nor is it new. Social psychologist William McDougall noted the parallel between emotions and colors in 1921, writing that "the color sensations present, like the emotions, an indefinitely great variety of qualities shading into one another by imperceptible gradients...." The first circumplex model was one developed by Brown University psychologist Harold Schlosberg in 1941, after he had asked research participants to judge the emotions posed in a standard set of pictures of facial expression. Schlosberg added the intensity dimension to his model. My own model was proposed in 1958, when I suggested eight basic bipolar emotions: *joy* versus *sorrow*, *anger* versus *fear*, *acceptance* versus *disgust* and *surprise* versus *expectancy*." *Ibid.* p. 349.