

Prise de vues virtuelles

Tourner un film avec un ordinateur

Théophile Farant

Mémoire de M2

Introduction	6
<i>Cinéma et animation : images du monde, images sans monde</i>	6
<i>Bouleversements numériques</i>	8
<i>Vers la prise de vues virtuelles</i>	12
PREMIÈRE PARTIE. Réalité et expérience de la prise de vues.	18
I. Réalité argentique	18
<i>De la représentation du monde à sa reproduction</i>	18
<i>La réalité sensible de la photographie</i>	22
<i>Prise de vues et mouvement</i>	25
<i>Le réalisme esthétique du cinéma : « de la reproduction à l'expression »</i>	28
<i>L'esthétique du réalisme : filmer la vie</i>	32
II. Réalité numérique	37
<i>Le paradigme mécaniste</i>	37
<i>Le paradigme numérique</i>	41
<i>Informatique graphique : des images aux pouvoirs nouveaux</i>	49
<i>Le réel du virtuel</i>	52
<i>De la reproduction à la simulation</i>	58
III. L'expérience de la prise de vues	66
<i>Contre la figuration naturaliste</i>	66
<i>Faire monde</i>	71

<i>Faire image</i>	77
<i>Un point de vue...</i>	80
<i>...Dans le monde</i>	84
DEUXIÈME PARTIE. Tourner avec un ordinateur	91
IV. Expérience de la prise de vues réelles numérique	91
<i>L'image-espace</i>	91
<i>L'image analogique</i>	95
<i>La libération numérique</i>	99
<i>L'augmentation de l'espace</i>	104
<i>L'hybridation de la prise de vues</i>	107
<i>Prise de volume : un point de vue de toutes parts</i>	110
V. Expérience de la synthèse 3D	115
<i>La synthèse d'images : entre animation et génération</i>	115
<i>Le rendu 3D : physically based et espace semi-persistent</i>	120
<i>Un plateau de tournage virtuel</i>	125
<i>Phénoménologie de la perception virtuelle</i>	131
<i>Être-au-monde virtuel</i>	139
VI. Expérience de la génération d'images	146
<i>Générer des films</i>	146
<i>Les modèles de génération</i>	150
<i>Pixel space et espace latent</i>	156
<i>Des images sans monde</i>	162

VII. Perspectives futures : cinéma de prise de vues virtuelles ou artificial live

action	177
<i>Générer des mondes...</i>	<i>177</i>
<i>...Peuplés d'agents</i>	<i>185</i>
<i>Par-delà prise de vues réelles et animation</i>	<i>196</i>
Études d'œuvres	202
<i>Nikita Diakur et la simulation du mouvement</i>	<i>202</i>
<i>Ian Cheng et la simulation de mondes</i>	<i>209</i>
Bibliographie	220

Introduction

Cinéma et animation : images du monde, images sans monde

Dans un ouvrage de référence, *The Language of New Media*¹, le théoricien Lev Manovich retrace l'histoire de l'image en mouvement qui trouve sa genèse dans les procédés d'animation. Ce sont d'abord les forains qui mettent au point l'image en mouvement, cette illusion du mouvement qui résulte de la succession rapide d'images fixes. Si les premiers dispositifs, à l'instar du praxinoscope, donnent à voir des animations de dessins ou de silhouettes de papier découpé, dès l'invention du cinématographe par les frères Lumière, le cinéma s'impose comme l'enregistrement du monde « réel » sur la pellicule. La prise de vues phagocyte alors le champ de l'image en mouvement ; elle en devient le genre dominant dont l'animation est reléguée au rang de sous-catégorie. Ainsi, contre un usage courant qui distingue entre « cinéma de prise de vues réelles » et « cinéma d'animation », il convient de rappeler qu'au sens premier du terme, le cinéma désigne la prise de vues réelles — l'« enregistrement photographique inaltéré d'évènements réels ayant eu lieu dans l'espace physique réel »² — à laquelle s'oppose donc l'animation.

Dès ses débuts, le cinéma se définit comme « l'art de l'index »³ : les images qu'il produit entretiennent un lien privilégié avec le réel dont elles sont les empreintes. Il hérite en cela de la révolution photographique qui a eu lieu quelques décennies plus tôt, dans la première moitié du XIXe. Pour la première fois de l'histoire, il est possible de déléguer la production d'images à une machine qui devient la garante d'une objectivité longtemps poursuivie. Dans le cadre de la figuration naturaliste qui, au moins depuis la Renaissance, sert d'idéal aux artistes occidentaux, l'image photographique représente la perfection en ce qu'elle met à distance la subjectivité humaine du processus de création. Dans la poursuite de l'imitation de la nature, les artistes du Quattrocento avaient déjà réalisé un grand pas en direction de l'objectivité avec la mise au point de la perspective et le recours aux sciences naturelles. Mais ils n'étaient pas parvenus à se défaire de la subjectivité inhérente à la représentation. C'est chose faite avec la photographie qui instaure le passage de la représentation à la reproduction et délivre du même coup la peinture de la figuration naturaliste, ouvrant la voie aux symbolismes de l'art moderne. L'image photographique, et à son tour l'image

¹ MANOVICH Lev, *The language of new media*, Cambridge, Mass., MIT Press, 2001.

² *Ibid.*, p. 249.

³ *Ibid.*, p. 250.

cinématographique, s'imposent comme parfaitement objectives, dépositaires d'une autorité directement héritée du réel. Qu'ils soient les représentants d'un cinéma-vérité qui pratique une prise de vues réelles inaltérée, fidèle en cela à l'école des frères Lumière, ou ceux du cinéma-illusion qui recourt à volonté aux techniques de trucages et d'effets spéciaux, descendants de George Méliès, les cinéastes s'accordent sur un point : le cinéma doit toujours se présenter comme l'enregistrement inaltéré du monde, la capture d'un « ça a été »¹ comme l'écrit Roland Barthes. Tourner un film c'est enregistrer une réalité déjà-là, qui nous précède et nous englobe. Au cours du XXe siècle, le cinéma améliore progressivement son réalisme jusqu'à devenir l'art immersif par excellence qui plonge ses spectateurs dans une réalité dont ils ont envie de croire/ont semblant de croire qu'elle s'est déroulée telle quelle.

À l'inverse, l'animation ne repose sur aucune suspension de l'incrédulité à propos de la nature de ses images. Chacun sait, lorsqu'il voit un film d'animation, qu'il a accès à un monde imaginaire, là où une large majorité des spectateurs de prise de vues réelles sont bien en peine de placer une limite nette entre les images qui ont été filmées et les manipulations visuelles des effets spéciaux. Et de fait, les frontières sont parfois floues. Que dire par exemple de l'animation sur *celluloid* ou de la *stop-motion*, technique qui consiste à animer des marionnettes ? Ces deux méthodes d'animation reposent bien sur l'enregistrement d'une réalité physique sur la surface de la pellicule. C'est que la dimension indicielle ne suffit pas à définir la prise de vues réelles. Ici, le terme anglophone de « live action », littéralement « action vivante », permet de mieux saisir la distinction entre cinéma et animation. La prise de vues réelles implique l'enregistrement d'une réalité vivante, douée d'un mouvement autonome, déjà-là. Par contraste, l'animation c'est l'art d'insuffler la vie ou, du moins, le mouvement, de manière suivie, en intervenant directement à chaque étape. Pour employer une métaphore religieuse, c'est une « création continuée » : l'animateur, comme le Dieu cartésien, doit contrôler chaque pose du mouvement, doit constamment remonter la grande horloge du monde, et même plus que ça puisqu'il n'y pas de mécanisme, il doit en faire tourner l'aiguille en la poussant du doigt. L'animation est un processus fastidieux, plus ou moins austère selon les techniques, qui peut donner l'impression de *se battre* contre ses personnages pour leur procurer un semblant de souffle. C'est un poids important sur les épaules de l'animateur dont la création est toute entière portée par la force de son imaginaire. De son côté, l'opérateur de prise de vues s'appuie sur sa perception, c'est plutôt un témoin-enregistreur du monde qui se

¹ BARTHES Roland, *La chambre claire: note sur la photographie*, Nachdruck., Paris, Gallimard Seuil, « Cahiers du cinéma », 2016, 192 p, p. 124.

déroule devant de lui : il n'a, en théorie, besoin que d'appuyer sur le bouton-enregistreur de la caméra pour que le film se constitue « tout seul ». Dans son ouvrage *Métamorphoses de l'intelligence*¹, la philosophe Catherine Malabou distingue entre deux formes d'automatisme : ce qui est « mouvement involontaire, qui n'a pas en soi-même son “âme” », et « ce qui se meut de soi-même, spontanément »². L'animation se trouve du côté du premier automatisme, celui du mouvement des automates, du côté de l'inerte à partir duquel elle s'efforce de produire l'illusion d'un mouvement autonome. Le *live action* correspond, lui, au second automatisme, celui du monde vivant. On saisit alors clairement la distinction entre cinéma et animation : l'un se situe du côté du réel, il repose sur la perception et l'enregistrement du monde, l'autre renvoie à une forme d'irréel, elle repose sur l'imaginaire ; les images du cinéma sont des images du monde, tandis que celles de l'animation sont des images sans monde. Mais ces catégories, très nettes tout au long du règne de l'argentique, sont brouillées avec l'arrivée du numérique, la perte de l'indicialité et l'avènement des images de synthèse.

Bouleversements numériques

La première interface graphique, *SketchPad*, est mise au point par Ivan Sutherland, père de l'informatique graphique et de la réalité virtuelle, en 1963. Celle-ci permet à ses utilisateurs, munis d'un stylet, de dessiner directement sur l'écran de l'ordinateur. C'est une petite révolution dans le sens où il s'agit là de la première image-interface, servant d'intermédiaire entre un humain et une machine. Mais *SketchPad* est encore bien loin des préoccupations du cinéma. La même année, sont tournés des films comme *Le mépris* de Jean-Luc Godard, *8½* de Fellini ou encore le démesuré *Cléopâtre* de Mankiewicz, et l'on voit mal le lien qu'aurait pu tracer ces réalisateurs entre leurs œuvres d'un art au sommet de sa maturité et la trouvaille amusante d'un jeune chercheur de cette nouvelle discipline qu'est alors l'informatique. Mais Sutherland rejoint bientôt l'université de l'Utah dont il développe le pôle d'informatique graphique qui deviendra central dans l'histoire des images numériques. Parmi ses étudiants, le pôle accueille Ed Catmull, futur fondateur de *Pixar*, qui réalise le premier film d'animation 3D en 1972, *A computer animated hand*. Encore une fois, pas de quoi inquiéter le cinéma car, malgré la prouesse technique qu'il représente, le « film » de Catmull est encore bien rudimentaire : une main de synthèse, calquée sur sa propre main, se contracte et se

¹ MALABOU Catherine, *Métamorphoses de l'intelligence: du QI à l'IA*, Paris, PUF, « Quadrige », 2021.

² *Ibid.*, p. 127.

rétracte pendant quelques minutes en tournant sur elle-même avant de pointer du doigt le spectateur. Mais la technologie fait son chemin, et c'est un camarade de Catmull, Fred Parke, qui réalise la première séquence d'animation 3D intégrée dans un long-métrage sorti au cinéma, *Futureworld* en 1976. La jonction est faite et, dès lors, les images de synthèse vont peu à peu s'intégrer aux effets spéciaux. D'abord minoritaires, ces nouvelles images prennent de plus en plus de place sur le grand écran. Les spectateurs des salles obscures découvrent *Tron*, produit par Walt Disney en 1982, qui impressionne par ses quelques vingt minutes d'animation 3D d'une esthétique futuriste, absolument inédite pour l'époque. Au cours de la décennie suivante, les images de synthèse se démocratisent et franchissent un nouveau palier avec *The abyss* de James Cameron. Sorti en 1989, le film marque les esprits pour être le premier à intégrer des images de synthèse de manière « invisible » au milieu des prises de vues réelles. Le studio *ILM*, fondé par Spielberg, qui en a réalisé les effets spéciaux, ouvre la voie au photoréalisme de la 3D. S'ensuivent de nombreux films au cours des années 1990 qui repoussent à chaque fois les limites techniques et artistiques : *La belle et la bête* de Walt Disney et *Terminator 2* de James Cameron, deux gros succès du *box-office* de 1991, puis *Jurassic Park* de Steven Spielberg en 1993 et, bien sûr, *Toy Story* de Pixar en 1995, premier long-métrage entièrement réalisé en animation 3D. Aujourd'hui, l'écrasante majorité des films en « prise de vues réelles » comportent des images de synthèse ; de manière évidente, dans les *blockbusters* à effets spéciaux où ce sont souvent les images en prise de vues réelles qui se font rares — le premier *Avatar*, tenu pour un film en « prise de vues réelles », comporte 117 minutes d'animation 3D sur les 160 de la durée totale — mais également de manière plus subtile, y compris dans les « films d'auteur », lorsqu'il s'agit de modifier discrètement un visage, les contours d'une silhouette ou, bien souvent, de réparer une inadvertance commise lors du tournage — peu de réalisateurs s'en privent. Au cours de vingt-cinq dernières années, les images de synthèse ont étendu lentement, mais sans faillir, leur empire sur le cinéma, remplaçant de manière indétectable une part de plus en plus conséquente des images capturées directement par la caméra. Qu'on se rende compte de l'ampleur du glissement ontologique de l'image cinématographique ! Toute sa réalité, toute sa force de conviction reposait, depuis son origine, sur cette dimension indicielle qui faisait d'elle une forme de prolongement de l'objet reproduit dont elle allait jusqu'à posséder l'ontologie. Le critique André Bazin écrivait : « L'image peut être floue, déformée, décolorée, sans valeur documentaire, elle procède par sa genèse de l'ontologie du modèle ; elle est le modèle. »¹. La grande nouveauté de l'image cinématographique dans l'histoire de l'art, et ce qui la séparait de l'animation, c'était

¹ BAZIN André, *Qu'est-ce que le cinéma ?*, Paris, Cerf, « 7e art », 1985, p. 14.

justement sa mise à distance de l'humain dans le processus de production de l'image. Si la subjectivité du réalisateur pouvait s'exprimer avant ou après, elle n'avait aucun moyen de s'immiscer au milieu, dans la chambre noire de la caméra : la genèse de l'image cinématographique n'impliquait que la lumière du monde et la pellicule, en contact direct. Or, l'image de synthèse réintroduit la subjectivité humaine au cœur du processus, plus proche en cela de la représentation picturale que de la photographie. De nouveau, l'objectivité des images est compromise, retombe dans l'ontologie fragile de la représentation sur laquelle plane toujours l'ombre suspecte de la subjectivité. Mais ce n'est pas tout, non seulement l'image de synthèse remet en cause le statut indiciel du cinéma, mais, en plus, elle se fait passer pour l'image photographique sans nous laisser aucune chance de la confondre. C'est que la photographie s'est imposée culturellement depuis la fin du XIXe comme le critère du vrai, l'étalon esthétique de l'image « réelle ». Or, les images capturées par des caméras virtuelles dans des univers en 3D sont trop parfaites pour appartenir ce réel-là. Alors on les tord, on les déforme, on simule de nombreux artefacts propres à la prise de vues réelles qui n'ont pas lieu d'être dans la fabrique des images de synthèse. En plus du grain reproduisant la texture si particulière de la pellicule, les infographistes rajoutent à l'envie des aberrations chromatiques, des distorsions de lentilles ou encore des traces de saleté sur le verre de l'objectif comme si un opérateur fantôme avait manipulé la caméra virtuelle de ses doigts gras. Dans cette poursuite du photoréalisme, l'image de synthèse devient finalement doublement suspecte : elle n'a ni l'objectivité de la photographie, ni l'honnêteté du symbolisme pictural qui s'assume. On comprend alors la réticence de certains réalisateurs à avouer l'usage extensif qu'ils en font. Christopher Nolan ou Ridley Scott, entre autres, ont déclaré à plusieurs reprises en interview que leurs films *Oppenheimer* et *Napoléon* s'étaient appuyés sur une utilisation minime des effets spéciaux numériques au profit d'effets spéciaux « pratiques », i.e. réalisés physiquement. Ces prises de position, critiquées par la communauté des artistes *VFX* dont certains avaient eux-mêmes travaillé sur les films en question, s'inscrivent dans un mouvement général de dénigration, ou au moins de minimisation, de la *CGI* (*Computer Generated Images*) au cinéma.¹ Les images de synthèse, trompeuses et traîtresses, apparaissent comme un procédé peu avouable qu'il vaut mieux dissimuler.

Et la difficulté ontologique ne s'arrête pas là : ce sont toutes les images numériques, y compris celles issues de la prise de vues réelles, dont le statut devient douteux. En 1969, est inventé le premier capteur numérique qui simule l'empreinte des photons sur la pellicule. D'abord moquées

¹ Lire à ce propos l'article « What do you mean, no CGI ? » du magazine *VFXV*: <https://www.vfxvoice.com/what-do-you-mean-no-cgi/>

pour leur mauvaise qualité, les caméras numériques finissent par s'imposer sur les tournages à partir des années 2000. Aujourd'hui, toutes les images capturées en prise de vues réelles sont de nature numérique, y compris les quelques-unes encore tournées sur pellicule qui n'échappent pas pour autant, en fin de chaîne, à un scan numérique. De ceci, résulte l'homogénéité ontologique des images à l'ère numérique. Quelle que soit leur origine — création de synthèse dans un logiciel en trois dimensions, prise de vues réelles ou encore dessin sur une tablette — toutes ces images partagent la même nature fondamentale : ce sont des grilles de *pixels* codées par des séquences de *bits*. Avec une conséquence simple : toute image est désormais manipulable et, comme le souligne Manovich, l'image en mouvement devient animation de *pixels*. Forte de cette ontologie déroutante, l'image numérique pose une question existentielle au cinéma qui se définissait jusque-là par cette « tentative de faire de l'art à partir d'une empreinte »¹. C'est cela, cette empreinte du monde, que remet en cause le cinéma numérique. On s'explique alors sans peine l'effacement de nombreux cinéastes face à cette rupture du contact direct entre la lumière et la pellicule, parfois vécue comme une perte du réel. Il est difficile de faire le deuil d'une objectivité si longtemps poursuivie et qu'on croyait enfin tenir avec la prise de vues argentique qui a constitué la caractéristique commune, sans doute la plus importante, de tous les films du XXe siècle :

Du point de vue d'un futur historien de la culture visuelle, les différences entre les films hollywoodiens classiques, les films d'art européens et les films d'avant-garde (à l'exception des films abstraits) peuvent sembler moins importantes que cette caractéristique commune : ils s'appuyaient tous sur des enregistrements de la réalité à l'aide d'objectifs photographiques.²

Le numérique rebat complètement les cartes et force le cinéma à se redéfinir. Manovich se plaît à souligner le cercle que décrit le cinéma au XXe siècle : parti de l'animation qu'il a vite reléguée au rang de sous-catégorie d'un art premier, celui de la prise de vues réelles, le voici qui y retourne, qui se plonge dans ses origines comme dans un bain de jouvence.

La construction manuelle d'images dans le cinéma numérique représente un retour aux pratiques pré-cinématographiques du XIXe siècle, lorsque les images étaient peintes et animées à la main. Au tournant du XXe siècle, le cinéma a délégué ces techniques manuelles à l'animation et s'est défini comme un moyen d'enregistrement. À l'ère du numérique, ces techniques redeviennent courantes

¹ L. Manovich, *The language of new media*, op. cit., p. 250. Traduction personnelle : « an attempt to make art out of a footprint. »

² *Ibid.*, p. 249. Traduction personnelle : « From the perspective of a future historian of visual culture, the differences between classical Hollywood films, European art films and avant-garde films (apart from abstract ones) may appear less significant than this common feature: that they relied on lens-based recordings of reality. »

dans le processus de réalisation cinématographique. Par conséquent, le cinéma ne peut plus être clairement distingué de l'animation. Il ne s'agit plus d'une technologie médiatique indexicale, mais plutôt d'un sous-genre de la peinture.¹

Avec l'irruption du numérique, le rapport de force s'est inversé : puisque tout n'est plus qu'animation de *pixels*, le cinéma en prise de vues réelles, tourné avec des caméras numériques, n'est plus à son tour qu'une sous-catégorie de l'animation. Manovich décrit la nouvelle identité du cinéma numérique comme hybride, alliage de prise de vues réelles, de peinture, d'opérations sur l'image, d'animation 2D et 3D. Ce nouveau régime ontologique nous oblige à remettre en cause « le pouvoir irrationnel de la photographie qui emporte notre croyance »² tel que le décrivait Bazin. Les images de synthèse ont enfoncé le premier coin, la calculabilité inhérente à la nature même de l'image numérique s'est chargée du second, l'IA générative finit d'emporter avec fracas l'édifice construit patiemment et à l'allure pourtant solide : il nous faut définitivement enterrer notre croyance dans l'objectivité des images photoréalistes. Le « ça a été » de la prise de vues réelles est soumis à une remise en cause sévère au profit des images sans monde de l'animation, invitant à une redéfinition du cinéma.

Vers la prise de vues virtuelles

En perdant leur indicialité, les images du cinéma perdent leur statut ontologique à part et leur autorité tirée d'un lien privilégié avec le réel. Surtout, les manipulations rendues possibles par le numérique réduisent l'ensemble de l'image en mouvement à de l'animation de *pixels*. C'est particulièrement flagrant dans les films composés en grande partie d'images de synthèse comme *Avatar* dont nous venons de citer l'exemple. Plus des deux tiers du film étant générés par ordinateur, on sent bien que sa classification en tant que « film de prise de vues réelles » pose problème. Il apparaît plus logique de le placer du côté de l'animation, tout comme l'ensemble des images de synthèse qui sont calculées à partir d'espaces virtuels sortis tout droit de l'imaginaire d'infographistes qui leur donnent vie à force de efforts laborieux et répétés. Pourtant, aussi logique soit-elle, cette catégorisation des images de synthèse en tant qu'animation semble manquer quelque

¹ *Ibid.*, p. 250. Traduction personnelle : « the manual construction of images in digital cinema represents a return to nineteenth century pre-cinematic practices, when images were hand-painted and hand-animated. At the turn of the twentieth century, cinema was to delegate these manual techniques to animation and define itself as a recording medium. As cinema enters the digital age, these techniques are again becoming the commonplace in the filmmaking process. Consequently, cinema can no longer be clearly distinguished from animation. It is no longer an indexical media technology but, rather, a sub-genre of painting. »

² A. Bazin, *Qu'est-ce que le cinéma ?*, op. cit., p. 14.

chose. De quel côté, par exemple classer les machinimas, ces films tournés dans des jeux vidéo ? Bien qu'ils soient entièrement constitués d'images de synthèses, il paraît difficile de parler de « films d'animation » à leur propos. Les réalisateurs de machinimas sont d'abord des joueurs qui se trouvent plongés dans des univers virtuels dont il semble bien qu'ils enregistrent la réalité. Le merveilleux documentaire *Knit's Island* réalisé par Ekiem Barbier, Quentin L'Helgoualc'h et Guilhem Caussecas constitue un exemple frappant d'équivocité. Tourné dans le jeu vidéo *DayZ*, les trois réalisateurs vont à la rencontre des autres joueurs pour les observer d'une manière presque anthropologique, pour échanger et éventuellement les interviewer. Le jeu qui met en scène un univers dystopique prenant place après un effondrement de l'ordre mondial s'appuie sur une mécanique survivaliste. Pour tourner leur film, les réalisateurs doivent donc d'abord s'assurer de la survie de leurs avatars, leurs doublures virtuelles, ce qui les amène à prendre en compte tout un tas de considérations très prosaïques : dormir, manger, boire, se déplacer, etc... Toujours, bien sûr, de manière virtuelle. Par ailleurs, aucune des images du film n'a été conçue directement par Ekiem Barbier et ses coréalisateurs, pas plus que les actions des personnages qu'ils rencontrent et derrière lesquels se cachent des joueurs humains face à leur PC. Tout est le fait du jeu, déjà-là, dans lequel ils se trouvent jetés comme dans un monde. Les réalisateurs parlent eux-mêmes de « tournage » à propos de la réalisation de *Knit's Island* : il ne fait aucun doute que les captures d'écran de ce monde virtuel dans lequel ils débarquent tiennent bien plus de la prise de vues que de l'animation. Pour autant, on peinerait également à classer le documentaire dans la catégorie de la « prise de vues réelles ». *Knit's Island* se trouve ainsi dans une sorte de *no man's land* conceptuel où ni la catégorie d'animation ni celle de prise de vues réelles ne semble s'appliquer.

Que dire également du travail du réalisateur d'animation 3D Nikita Diakur qui construit les décors de ses films en *camera mapping* (technique consistant à créer des objets virtuels 3D à partir de photographies) et s'appuie sur des simulations physiques, notamment celle de la gravité, pour animer ses personnages. Loin d'une création continuée, Diakur se contente au contraire de « lancer » des simulations informatiques et d'enregistrer leur déroulement. Et dans son court-métrage *Backflip*¹, Diakur va encore plus loin en mettant en scène une doublure numérique de lui-même qui apprend à réaliser un salto arrière dans un environnement virtuel 3D en s'appuyant sur du *machine learning*. Le film est le résultat d'un montage des meilleurs « prises » (les captures d'écran) du tournage virtuel qui a duré plusieurs semaines. Ici, le personnage principal est simulé physiquement et « interprété » par une intelligence artificielle qui, à force d'itérations et de

¹ Disponible sur Youtube : <https://youtu.be/1zUJzzRu-xs?si=KMm8f63CZECLmahK>

nombreuses chutes, parvient enfin à réatterrir sur ses « pieds » après avoir effectué un salto arrière. Difficile alors de ne pas éprouver une forme d'empathie étrange pour cet avatar qui paraît si vivant. Et de la même manière, l'artiste Ian Cheng conçoit des écosystèmes virtuels soutenus par des intelligences artificielles. Dans ces œuvres qui sont exposées sur des écrans géants, des créatures évoluent de manière autonome en interagissant avec leur environnement, procurant aux spectateurs l'impression de regarder dans un vivarium, comme si l'écran n'était qu'une vitre les séparant d'un monde virtuel. Et en effet, pour décrire sa pratique artistique, Cheng parle de « World Building » : construction de mondes. Chez Diakur ou Cheng, qui pourtant travaillent avec des logiciels d'animation 3D classiques, on ressent clairement l'ambition, œuvre après œuvre, de tendre vers la simulation de mouvements autonomes, d'automatismes qui tiennent plus du vivant que des automates. Si leurs films sont toujours classés du côté de l'animation, leur pratique évoque en revanche la prise de vues dans des monde virtuels.

Enfin, sans aller aussi loin du côté de l'animation expérimentale, de nombreuses technologies de création d'images de synthèse semblent brouiller les pistes. L'analogie de la prise de vues est particulièrement pertinente pour illustrer le fonctionnement des moteurs de rendu 3D qui constituent le mode paradigmatique de synthèse d'images. Ce sont ni plus ni moins que des simulations informatiques de la prise de vues : pour « rendre » une image, il faut placer une caméra virtuelle, dont on règle les paramètres optiques (ouverture, focale, sensibilité *ISO*, etc...), et des spots lumineux pour éclairer la scène selon les lois physiques de diffusion de la lumière ; une fois que tout est en place – que les objets virtuels sont convenablement disposés dans l'espace tridimensionnel et que l'éclairage nous convient – il suffit alors d'appuyer sur le bouton de rendu qui, comme un déclencheur virtuel, entraîne la création automatique d'une image qui n'est au fond pas autre chose que l'enregistrement d'un « ça a été » virtuel. Et, si les moteurs de rendu utilisés par l'industrie de l'animation peuvent mettre plusieurs heures, voire plusieurs jours, pour le calcul d'une seule image, les moteurs de jeu vidéo en temps réel rendent les images dans l'instant qui précède leur affichage. Dès lors, plus rien ne s'oppose à l'intégration d'images de synthèse dans un flux d'images réelles au moment même où elles sont tournées. Les moteurs de rendu *realtime* s'invitent sur les plateaux de tournage. La *motion capture* en temps réel permet, par exemple, de retranscrire en direct les mouvements d'un acteur revêtu d'une combinaison de capteurs sur une créature virtuelle. De son côté, la prévisualisation en temps réel (*On-Set Previz*) permet de visualiser directement dans le retour caméra le *compositing* des images de synthèse et des images capturées, tandis que les *LED Walls*, vastes écrans sur lesquels sont affichés des décors virtuels reliés aux mouvements de la caméra physique, remplacent les fonds verts devant lesquels jouaient

les acteurs. Le rendu en temps réel permet ainsi l'intégration des images de synthèse au moment même du tournage, et non plus de manière différée pendant la phase de *post-production*. L'ensemble de ces pratiques sont regroupées sous l'appellation de *virtual cinematography* qui renvoie à l'application d'une connaissance pratique des caméras et des lumières physiques aux caméras et lumières virtuelles. En effet, les chefs opérateurs habitués à la prise de vues réelles se forment peu à peu aux techniques de synthèse d'images pour répondre aux besoins nouveaux des tournages contemporains. Mais dans cet échange, est-ce la prise de vues réelles qui tend vers l'animation ou l'animation qui tend vers la prise de vues réelles ?

On voit bien, à travers ces différents exemples, que la distinction entre prise de vues réelles et animation n'est pas aussi rigide qu'elle le paraissait au premier abord, surtout elle n'est plus suffisante. Toutes ces technologies de synthèse d'images tendent vers un automatisme de plus en plus souple et autonome à mesure des progrès techniques. C'est notamment la promesse que porte l'IA générative qui offre déjà des résultats impressionnants dans la génération d'images et de vidéo. À partir d'un simple *prompt*, il est possible de générer de très nombreuses variations, ouvrant la voie à un nouveau type de création automatique d'image en mouvement. Et au-delà de la vidéo, la recherche en intelligence artificielle s'intéresse déjà à l'étape suivante : les « worlds models », capables de simuler des mondes, compris comme des systèmes complets, et des agents pour les peupler. À mesure que l'automatisme des simulations informatiques gagne en autonomie, les espaces virtuels semblent se transformer en véritables mondes. Face à de telles perspectives, la thèse de Manovich selon laquelle à l'ère numérique toute les images se valent, aussi pertinente soit-elle, nous semble insuffisante. Elle se place du côté des images finies, des images objets. Et en effet, de ce point de vue toutes les images sont bien des grilles de *pixels*. Mais c'est manquer une dimension fondamentale, sans doute la plus importante dans l'ontologie des images : la façon dont elles sont faites. Ce qui différencie la prise de vues de l'animation, bien plus qu'un caractère indiciel ou non indiciel qui n'a plus de sens à l'époque numérique – et en cela nous sommes d'accord avec Manovich – c'est l'expérience de création. Pour la prise de vues, cette expérience c'est la perception d'un monde dans lequel on tourne – vivant, toujours au-devant de nous, apparaître permanent, totalité inachevée, lieu d'intersubjectivité. Pour l'animation, c'est l'expérience d'un imaginaire qui repose tout entier sur l'esprit de l'animateur. Dès lors, contre Manovich, nous éprouvons l'intuition que la distinction entre les catégories de cinéma et d'animation a encore un sens à l'ère numérique. Mais à une condition : que la prise de vues sur laquelle s'appuie la définition classique du cinéma s'élargisse pour intégrer la prise de vues virtuelles conçue comme l'enregistrement de mondes virtuels. On l'a vu, ni les catégories

d'animation, ni celle de prise de vues réelles ne suffisent pour désigner les potentialités des simulations informatiques. Face à l'émergence de films tournés dans des mondes virtuels à l'instar de *Knit's Island*, la définition d'une nouvelle catégorie qui serait la prise de vues virtuelles fait pleinement sens. Il y aurait donc, d'un côté, les images de mondes, qu'ils soient physiques ou virtuels, reposant sur la perception et appartenant à la prise de vues en général ; et de l'autre, les images sans monde, reposant sur l'imaginaire et relevant de l'animation. Bien sûr ces deux catégories sont poreuses et, sans doute, rares sont les images pures, se situant exclusivement d'un côté ou de l'autre. Mais cette séparation possède la vertu de rendre possible une approche phénoménologique de la création d'images de synthèse. Ce serait alors l'occasion de réaliser un nouveau type de cinéma qui embrasserait toutes les potentialités du numérique, un cinéma qui ferait le pari que les simulations informatiques peuvent donner lieu à des mondes au sens plein, ouvrant la voie à une manière de *tourner des films avec un ordinateur* qui procurerait un engagement aussi riche que le tournage d'un film en prise de vues réelles.

De nombreuses questions s'ouvrent alors : Quelle est l'ontologie des images cinématographiques argentiques ? Et celle des images numériques ? Qu'est-ce qui les sépare et qu'est-ce qui les rapproche ? Quelle est l'expérience de la prise de vues ? À quelles expériences de création donnent lieu les différentes techniques de « tournage avec un ordinateur » ? Les méthodes de synthèse d'images permettent-elles une forme de perception virtuelle, condition nécessaire pour la prise de vues ? Les simulations informatiques seront-elles en mesure de dépasser l'automatisme des machines pour atteindre l'automatisme plastique du vivant ? Pourront-elles donner lieu à des mondes virtuels au sens plein ? Enfin, la prise de vues virtuelles est-elle possible ?

Nous divisons notre enquête en deux grandes parties. La première, *Réalité et expérience de la prise de vues*, composée de trois chapitres, pose les bases ontologiques et phénoménologiques qui nous serviront ensuite. La seconde, *Tourner avec un ordinateur*, composée de quatre chapitres, étudie l'expérience concrète de la réalisation de films avec différentes technologies de « prise de vues virtuelles » – prise de vues réelles numérique, synthèse 3D, génération de vidéo par IA et *world models*.

Le premier chapitre s'intéresse au lien privilégié qu'entretient le cinéma argentique avec la réalité du fait de son caractère indiciel que nous replaçons dans une histoire de la figuration naturaliste. Nous retraçons, en particulier, les différentes étapes de la représentation picturale à la reproduction photographique, et puis de la reproduction photographique à l'expression cinématographique avec le développement d'une grammaire du cinéma.

Le second chapitre s'intéresse à nature de l'informatique, du virtuel et de ses images. Nous voyons comment le virtuel numérique marque le passage de la reproduction à la simulation, poursuivant en cela l'entreprise de la figuration naturaliste.

Le troisième chapitre abandonne les investigations ontologiques qui se placent du point de vue de l'image figée pour s'intéresser à l'images en train de se faire. Contre l'idéal d'objectivité de la figuration naturaliste, les images sont toujours le produit d'une expérience vécue. Dans cette optique, nous développons une approche phénoménologique de la prise de vues.

Le quatrième chapitre marque l'entrée dans le domaine du concret. Nous nous intéressons à un premier cas de « tournage avec un ordinateur », la prise de vues réelles numérique. Cette révolution technologique change drastiquement l'expérience du tournage. Sans compter sur sa dimension numérique qui ouvre la voie à toutes les manipulations, entraînant une hybridation des images.

Le cinquième chapitre poursuit dans le thème de l'hybridation en s'intéressant à l'expérience de la synthèse 3D. Les moteurs de rendu sont de véritables simulations de la prise de vues réelles et du monde physique. Surtout, ils donnent lieu à des espaces tridimensionnels qui permettent une perception virtuelle, ouvrant la voie à une forme de prise de vues virtuelles. Mais les moteurs de rendu se heurtent à leur automatisme déterministe et à un manque d'autonomie.

Le sixième chapitre explore le fonctionnement de la génération d'images et de vidéos par intelligence artificielle. Les modèles de diffusion ne permettent qu'une expérience spatiale très pauvre, tenant plus de l'imaginaire que de la perception virtuelle. En revanche, ils déploient un automatisme très souple qui permet un processus créatif assez riche et surtout prometteur.

Enfin, le septième chapitre explore les promesses d'autonomie entrevues au chapitre précédent. Nous évoquons une nouvelle architecture de modèle, les *worlds models*, à même des générer des mondes structurés en temps réel, et la mise au point d'agent virtuels incorporés pour les peupler. Le changement de paradigme introduit par ces technologies, avec le passage d'une IA symbolique à une IA incarnée, semble annoncer un dépassement des difficultés rencontrées jusqu'à là : le manque d'autonomie des moteurs 3D et l'absence de perception virtuelle de la génération de vidéo. La prise de vues virtuelles se dessine à l'horizon.

PREMIÈRE PARTIE. Réalité et expérience de la prise de vues.

I. Réalité argentique

De la représentation du monde à sa reproduction

Les plus anciennes traces d'une activité artistique — les peintures rupestres pour certaines datées d'environ - 30 000 — sont déjà des représentations du monde. Les spécialistes insistent sur leur réalisme, la façon dont nos ancêtres ont su utiliser les aspérités de la roche, ses renflements ou ses creux pour donner du relief au corps musclé d'un bison ou du mouvement à la trajectoire d'une lance. Dès le début, il semble que l'art ait été figuratif, concentré sur une imitation du réel. Le critique de cinéma André Bazin dresse une analogie entre la pratique de l'embaumement et la pratique artistique. L'art, selon Bazin, a comme but premier d'embaumer le temps, de « sauver l'être par l'apparence »¹. Ainsi, cette pratique est-elle hautement humaine, elle est le combat de l'humanité contre sa condition, une espèce de semi-victoire sur la mort qui n'est elle-même que « la victoire du temps »². Mais quelle est cette « apparence » dont parle Bazin ? C'est celle de l'enveloppe physique du monde, l'histoire des arts plastiques étant, selon le critique, « essentiellement celle la ressemblance ou, si l'on veut, du réalisme. »³. C'est que Bazin parle depuis une société et une époque précise, celles de la Modernité qui, comme toutes les sociétés et toutes les époques, possèdent une manière de concevoir le monde qui leur est propre. L'anthropologue Philippe Descola parle à ce propos d'« ontologies » pour désigner ces systèmes qui, non seulement, façonnent la manière dont une société conçoit le monde, mais aussi le perçoit et le figure. En l'occurrence, l'ontologie de la Modernité, celle qui nous intéresse car c'est cette société qui inventa la photographie, est, toujours selon Descola, celle du « naturalisme » qui pose une discontinuité des êtres en intériorité (l'humain doté d'un esprit est à part de la nature) mais une continuité en extériorité (la matière est partagée de manière homogène). Ainsi, l'idéal de figuration reposant sur l'ontologie naturaliste est-il bien celui de la ressemblance physique évoquée plus haut par Bazin. Il faut bien avouer que la figuration naturaliste porte en elle une force de conviction qui anime l'histoire de l'art, au moins occidental, bien longtemps avant même l'avènement de la

¹ A. Bazin, *Qu'est-ce que le cinéma ?*, op. cit., p. 8.

² *Ibid.*, p. 9

³ *Ibid.*, p. 10

Modernité, et qui tient à un argument simple : la ressemblance dote les images d'une formidable puissance d'agir. Descola en rend compte dans son livre *Les formes du visible* au sous-titre évocateur, « Une anthropologie de la figuration » :

Quelle façon plus immédiate, en effet, de conférer une puissance d'agir aux images que de les faire ressemblantes, si ressemblantes qu'elles en viennent à tromper l'œil en donnant l'illusion d'être la même réalité même et non son substitut ? Après tout, le mythe fondateur de la figuration naturaliste n'est-il pas le célèbre concours d'imitation qui, deux mille ans avant l'*ars nova*, voit s'opposer Zeuxis et Parrhasius : tandis que la grappe de raisin dépeinte par Zeuxis attire des oiseaux qui viennent le becqueter, le voile couvrant la peinture de Parrhasius dupe Zeuxis, qui, voulant l'écarter, s'aperçoit qu'il est peint. L'un avait leurré des oiseaux gourmands, l'autre avait abusé le regard expert d'un peintre.¹

Si la tentation de la figuration naturaliste est déjà présente dans les idéaux esthétiques de l'Antiquité, c'est à la Renaissance, moment charnière qui, justement, fait basculer l'Occident dans le naturalisme, qu'elle se concrétise. La poursuite de la ressemblance physique lie son histoire à celle des progrès techniques, chacun d'entre eux constituant un perfectionnement dans l'imitation du monde. Or le moment de la Renaissance se révèle particulièrement foisonnant en innovations. Il semble que la société artistique occidentale de l'époque se soit arc-boutée, d'un seul bloc puissant et compact, contre l'irréductible subjectivité de la représentation. Au moment où sont posés les fondements de la science moderne expérimentale, il n'est pas étonnant de voir les peintres et les sculpteurs occidentaux lorgner du côté de la vérité scientifique, attirés par « un art neuf, plus fidèle à la nature que tout ce qui avait été fait auparavant [...] un art plus vivant et plus direct »². L'historien de l'art Ernst Gombrich décrit comment le XVe siècle voit la création artistique occidentale se muer en une « conquête de la réalité »³, car ce qui est visé, c'est bien le désir de « remplacer le monde extérieur par son double »⁴. L'un des progrès techniques les plus marquants est sans doute l'invention de la perspective qui constitue un exemple saisissant de cette volonté artistique de toucher à l'absolu. C'est en particulier à l'utilisation du procédé de la « camera obscura », dont les principes sont déjà connus depuis Aristote, que l'on doit la mise au point de

¹ 17/08/2025 23:22:00

² GOMBRICH Ernst Hans, *Histoire de l'art*, Nouv. éd., Paris, Flammarion, 1994, p. 17.

³ *Ibid.*

⁴ A. Bazin, *Qu'est-ce que le cinéma ?*, *op. cit.*, p. 11.

cette nouvelle représentation du monde. Leonard De Vinci et Giovanni Battista della Porta, entre autres, améliorent le procédé ancien en le dotant d'une lentille et en font ainsi un outil pour le dessin, permettant la retranscription en deux dimensions de l'espace tridimensionnel. Cette technique témoigne de « l'idéal indiciel »¹ qui anime la peinture européenne. Et l'invention de la perspective s'inscrit dans cette dynamique nouvelle, mêlant art, science et technique, qui s'appuie sur la mise en application des principes de l'empirisme, et, de ce fait, mène à une révolution dans l'imitation de la nature que décrit Gombrich :

Les maîtres florentins de l'entourage de Brunelleschi, ont élaboré toute une méthode pour rendre la nature avec une précision quasi scientifique. Après avoir établi une armature de lignes perspectives, ils y édifiaient le corps humain, soutenus en cette entreprise par leur connaissance de l'anatomie et des lois du raccourci.²

Les tableaux de la Renaissance deviennent comme « le miroir du monde visible »³. Les artistes se mettent à observer la nature selon les règles de l'observation scientifique : ils assistent aux séances de dissection des facultés de médecine pour mieux comprendre l'anatomie, ils observent le comportement de la lumière et se livrent à des études géométriques concernant la perspective. Tous semblent vouloir répondre à Platon qui critiquait l'art trompeur qu'est la peinture, incapable de produire autre chose qu'une vulgaire imitation sans jamais atteindre à l'essence des choses. Au XVe siècle, la peinture fait presque oublier sa dimension imitatrice tant elle semble donner à voir la réalité elle-même : « pour la première fois dans l'Histoire, l'artiste était un parfait œil-enregistreur. »⁴, écrit Gombrich. C'est donc par l'alliance de l'art, de la science et de la technique que la Renaissance marque un tournant dans l'imitation humaine de la nature avec l'avènement de la figuration naturaliste, érigée en canon esthétique.

À côté de ces ambitions artistiques, les scientifiques ne sont pas en reste. Eux aussi, se lancent à la conquête de la nature avec la méthode empirique, et ses expériences, qui constituent également une forme de représentation esthétique. C'est sur ce nouveau concept d'« expérience », fondé sur l'observation et abandonnant tout idéalisme, que repose la science moderne. Le

¹ DESCOLA Philippe, *Les formes du visible, op. cit.*, p. 520

² E.H. Gombrich, *Histoire de l'art, op. cit.*, p. 178.

³ *Ibid.*

⁴ *Ibid.*

philosophe Bruno Latour, en se basant sur les travaux des historiens Steven Shapin et Simon Schaffer, attribue au scientifique irlandais Robert Boyle la parenté de l'empirisme moderne avec la réalisation de la première expérience scientifique, celle de la pompe à air mettant en évidence l'existence du vide :

Plutôt que sur la logique, les mathématiques ou la rhétorique, Boyle se fonde sur une métaphore para-juridique : des témoins crédibles, fortunés et de bonne foi rassemblés autour de la scène de l'action peuvent attester de l'existence d'un fait, the matter of fact, même s'ils n'en connaissent pas la véritable nature. Boyle invente ainsi le style empirique que nous utilisons encore aujourd'hui.¹

L'expérience du savant est une reproduction phénoménale du monde qui vise à rendre sensible des phénomènes naturels. Ainsi la science expérimentale, finalement proche de l'art en cela, se livre à une imitation de la nature : elle fabrique des faits, et c'est tout l'enjeu de la démonstration du constructivisme latourien. Le philosophe Gaston Bachelard va même jusqu'à parler de « phénoménotechnique » pour désigner la fabrication de phénomènes par le biais des instruments :

La division classique qui séparait la théorie de son application ignorait cette nécessité d'incorporer les conditions d'application dans l'essence même de la théorie. [...] C'est alors qu'on s'aperçoit que la science réalise ses objets, sans jamais les trouver tout faits. La phénoménotechnique étend la phénoménologie. Un concept est devenu scientifique dans la proportion où il est devenu technique, où il est accompagné d'une technique de réalisation.²

Avec ce terme de phénoménotechnique, Bachelard montre à quel point la science moderne est inséparable de la technique qui l'aide dans sa reproduction de la nature. Les instruments donnent un corps aux faits scientifiques, ils les dotent d'une phénoménalité. De ce point de vue, la science fabrique des faits en livrant une reproduction esthétique de la nature, de la chose-en-soi, par le biais d'expériences instrumentées : un microscope ou un écran donnent à voir des réalités phénoménales qui n'existeraient pas sans ces objets. Ainsi, le sort des images se retrouve-t-il lié, non pas seulement à celui de l'art, mais également à celui de la science moderne dont l'objectivité nouvelle doit bien trouver un support de présentation à la hauteur de sa Vérité. Descola note combien les scientifiques,

¹ LATOUR Bruno, *Nous n'avons jamais été modernes: essai d'anthropologie symétrique*, Nachdr., Paris, Editions La Découverte [u.a.], 2010, 206 p, p. 29.

² BACHELARD Gaston, *La formation de l'esprit scientifique*, 16. éd., Paris, Vrin, « Bibliothèques des textes philosophiques », 1999, 256 p, p. 61.

à l'instar des cartographes, s'intéressent de près aux mouvements de l'art vers la figuration naturaliste, de la mise à distance de la subjectivité dans la production des images. Trois siècles après la découverte fondatrice de la perspective, la photographie semble enfin concrétiser le rêve de De Vinci en révélant tout le potentiel de la camera obscura. Comme l'écrit Bazin, « la chambre noire de Vinci préfigurait celle de Niepce »¹. La mise au point de plaques photosensibles par l'ingénieur Nicéphore Niépce permet de fixer les images évanescentes que la camera obscura produisait déjà depuis des siècles. Après le premier procédé de Niépce, inventé en 1824, puis son amélioration par Louis Daguerre, peintre de formation, le daguerréotype est présenté en 1839 à l'Académie des Sciences, consacrant officiellement l'invention de la photographie. Il est significatif que cette révolution technique ait vu le jour grâce au travail combiné d'un scientifique et d'un artiste. Daguerre, au début de sa carrière, avait justement rencontré un certain succès avec l'invention du diorama, une technique de peinture en trompe l'œil, sur des plaques vitrées, animées par des jeux d'éclairage participant d'un plus grand réalisme. Au regard de cette histoire personnelle, les travaux photographiques de Daguerre apparaissent comme une poursuite de son entreprise artistique. Et c'est toute la photographie qui semble s'inscrire dans une histoire de la mimésis dont elle est un tournant majeur, entre continuité et rupture. Continuité, car, on l'a vu, cette invention ne fait que rejouer l'alliance déjà ancienne de l'art et de la technique au service de la figuration naturaliste. Rupture, car la photographie introduit une manière absolument nouvelle de concevoir des images, profondément différente de la représentation en ce qu'elle met désormais l'humain à distance : la prise de vues.

La réalité sensible de la photographie

Au premier abord, la prise de vues semble définie par son rapport immédiat, direct au monde qu'elle « capture ». La représentation, à l'inverse, implique une rupture du lien entre le monde et l'image qui le représente : quelque part dans la chaîne, il y a une interprétation qui rompt le contact, brouille les pistes et transforme la représentation en une approximation, plus ou moins rêvée, du monde. Le peintre qui reproduit un paysage l'observe du mieux qu'il peut pour ensuite le retracer le plus fidèlement possible sur la toile. Dans un tel processus, l'interprétation intervient donc à plusieurs niveaux. D'abord au niveau de la perception : la fiabilité de l'appareil perceptif humain a fait, et continue de faire, l'objet d'âpres débats d'ordre philosophique, psychologique ou scientifique. À quel point pouvons-nous nous fier à nos perceptions ? On connaît l'exemple

¹ A. Bazin, *Qu'est-ce que le cinéma ?*, op. cit., p. 10.

classique de la couleur qui, n'existant pas dans le monde extérieur, n'est présente dans notre perception que grâce aux bâtonnets de notre rétine. Mais le débat ne s'arrête pas là et s'étend à l'ensemble des objets et des qualités perçues. La perception peut-elle être divisée, comme le faisait Locke, entre qualités premières, indépendantes de l'esprit — donc perçues telles qu'elles existent dans le monde — et qualités secondaires, façonnées par l'esprit ? Dans une version plus actuelle du débat, percevons-nous seulement des *low-level properties* comme la couleur et le mouvement, ou avons-nous accès également à des *high-level properties* comme les objets naturels (un pin), les artefacts (une table) ou encore les qualités morales (être bon) ? Cette grande question de la fiabilité de nos perceptions s'apparente au problème de la pénétration cognitive : notre perception est-elle influencée par notre cognition ? Toutes ces questions demeureront sans doute ouvertes pour longtemps et nous n'essaierons pas de trancher ici le débat. Nous pouvons, en revanche, nous accorder, avec une grande majorité des observateurs, sur le rejet du réalisme naïf qui postule la fiabilité parfaite de l'appareil perceptif. Il est certain que nous ne percevons pas exactement le monde tel qu'il est, et que notre perception est, en ce sens, trompeuse. Être un grand artiste, c'est d'abord être un grand observateur, mais l'on voit bien ici que le plus grand des artistes ne sera jamais le « parfait œil-enregistreur » que décrivait Gombrich. Et quand bien même ce serait le cas, quand bien même les perceptions d'un artiste seraient absolument conformes au monde extérieur, nous ne serions encore qu'à la moitié du chemin : il resterait, pour cet artiste, à retranscrire de la manière la plus fiable, sans aucune altération, le contenu de sa perception dans son œuvre. Pour revenir à notre peintre de paysage, une fois que celui-ci a observé son modèle — avec plus ou moins de distorsion, plus ou moins de pénétration cognitive, mais certainement pas de manière exacte par rapport à la réalité objective — encore lui faut-il tracer sur sa toile les traits qui correspondent le mieux à sa perception, choisir les tons les plus fidèles à ceux qu'il a en tête. Il y a donc un exercice de sincérité chez l'artiste qui essaie de rester au plus près de ses sensations, mais cette sincérité est toujours une interprétation. Le résultat ne peut donc constituer qu'une nouvelle réalité, plus ou moins proche esthétiquement du modèle d'inspiration, mais nécessairement différente. En fin de compte, le modèle d'origine n'est plus que cela : une inspiration que le talent de l'artiste figuratif aura su garder plus ou moins proche. Cette dimension interprétative de la représentation a pour conséquence de rendre les images suspectes quant à leur objectivité. Comme l'a écrit Bazin, « si habile que fût le peintre, son œuvre était toujours hypothéquée par une subjectivité inévitable. Un doute subsistait sur l'image à cause de la présence de l'homme. »¹

¹ *Ibid.*, p. 12.

La photographie surgit dans ce combat historique de l'art figuratif pour se tenir au plus près de son modèle. Elle coupe court au problème de la double interprétation auquel se heurte la représentation. La prise de vues indique un rapport direct, et surtout, objectif, au modèle. L'appareil photographique est comme un nouvel appareil perceptif, plus fiable, échappant à toute pénétration cognitive, capable d'imprimer directement le résultat de sa perception sans avoir à passer par une nouvelle étape d'interprétation — comme si le peintre pouvait projeter sur sa toile l'image mentale qui résulte de son observation. L'image photographique diffère ontologiquement de l'image picturale du fait de son statut indiciel : c'est une empreinte du monde, là où la peinture n'est qu'une icône entretenant une relation de ressemblance, toujours douteuse, avec son modèle. Bien sûr, l'image indicielle n'est pas complètement nouvelle — les mains préhistoriques constituent même le point de départ de l'histoire de l'art — mais, avec la photographie, il devient possible de reproduire le monde de manière systématique et mécanique. L'image photographique apparaît à ses contemporains comme celle de la vérité du fait même de sa mise à distance de l'homme :

L'originalité de la photographie par rapport à la peinture réside donc dans son objectivité essentielle. Aussi bien, le groupe de lentilles qui constitue l'œil photographique substitué à l'œil humain s'appelle-t-il précisément « l'objectif ». Pour la première fois, entre l'objet initial et sa représentation, rien ne s'interpose qu'un autre objet. [...] Tous les arts sont fondés sur la présence de l'homme ; dans la seule photographie nous jouissons de son absence.¹

Les débuts de l'appareil photographique se font cependant un plus balbutiant que l'enthousiasme de Bazin : il faut plusieurs jours de pose à Nicéphore Niépce pour obtenir le *Point de Vue du Gras*, la première photographie recensée. Ce temps long n'est pas si éloigné de celui du peintre et favorise les distorsions perceptives (mouvements possibles de l'appareil, métamorphoses de l'environnement, changements lumineux...). Mais très vite, le temps de pose se réduit, notamment grâce au travail de Daguerre, et de nombreuses améliorations techniques s'ajoutent, emportant largement l'adhésion des contemporains, en particulier celle des scientifiques qui saisissent le potentiel du daguerréotype comme appareil d'enregistrement doté d'une objectivité à la hauteur de celle que requiert l'observation empirique. Ce qui attire les savants dans la photographie, « c'est la possibilité de s'en servir comme d'un instrument d'investigation engendrant des images libérées de

¹ *Ibid.*, p. 13.

l'interprétation humaine, donc véritablement objectives. »¹. Descola relève en particulier l'enthousiasme du botaniste Pierre-Jean François Turpin en 1839 pour qui les daguerréotypes appartiennent « presque entièrement à la nature ». C'est que l'image photographique, par sa dimension indicielle, est bien plus qu'une image, c'est une empreinte voire l'objet lui-même. De ce fait, elle acquiert très vite une autorité qui est celle du réel. On la préfère désormais à la peinture pour les planches scientifiques, pour les documents administratifs ou les actualités, car elle fait figure de preuve. Même si la photographie implique toujours une dimension interprétative dans les choix de mise en scène et de cadrage, l'interprétation est reléguée à la périphérie du processus de création, avant ou après l'enregistrement de l'image, mais certainement pas au milieu. Pour Barthes, la photographie est un « ça a été »², ce qui la dote d'une force ontologique indiscutable qu'elle tire de son modèle et avec lequel on la confond :

Quelles que soient les objections de notre esprit critique nous sommes forcés de croire à l'existence de l'objet représenté, effectivement re-présenté, c'est-à-dire rendu présent dans le temps et dans l'espace. La photographie bénéficie d'un transfert de réalité de la chose sur sa reproduction. »³

La photographie s'impose au XIXe comme l'image de la vérité au point qu'elle deviendra le critère du réalisme avec lequel nous jugeons encore aujourd'hui. Du même coup, elle délivre la peinture de sa poursuite de la ressemblance, ouvrant la voie au symbolisme de l'art moderne. Enfin, la camera obscura a révélé son plein potentiel pour permettre la reproduction des trois premières dimensions du réel, celles de l'espace, achevant en cela les travaux annonciateurs des peintres de la Renaissance sur la perspective. Mais concernant la quatrième dimension, elle se heurte aux mêmes difficultés que la peinture : le temps lui échappe encore.

Prise de vues et mouvement

Dès son invention, la photographie est un enregistrement de mouvement. C'est particulièrement vrai pour les premières images qui nécessitent des poses longues de plusieurs jours, et constituent ainsi l'empreinte de longues périodes temporelles. Une photographie prise sur

¹ DESCOLA Philippe, *Les formes du visible: une anthropologie de la figuration*, Paris, Éditions Points, « Points Essais », 2023, p. 521.

² R. Barthes, *La chambre claire*, op. cit., p. 124.

³ A. Bazin, *Qu'est-ce que le cinéma ?*, op. cit., p. 14.

deux ou trois jours, comme le *Point de vue du gras*, enregistre le bruissement des feuillages, les oscillations du soleil, le glissement des nuages... Cette seule image est un concentré de temps, compact et continu. Mais ces vues du monde ne sont pas en cela tellement différentes des représentations picturales dont les « temps de pose » s'étalent sur des jours entiers. Un peintre de l'époque qui aurait voulu réaliser le point de vue du Gras, aurait sans doute mis à peu près le même temps pour un résultat pas moins réaliste. Ces poses longues ne satisfont pas les photographes qui cherchent, au contraire, à capturer le mouvement dans sa plus grande fugacité. Pour cela, il faut considérablement réduire le temps d'exposition des plaques photosensibles. À mesure des progrès techniques, les appareils permettent des prises de plus en plus courtes, ramenant le temps d'exposition sous la seconde, puis le dixième de seconde ; alors, il devient possible de parler d'« instantanés ». Soudain, la photographie est en mesure de fixer les gestes les plus furtifs, les mouvements trop rapides pour être saisis, par l'œil, autrement que dissimulés sous un épais flou de vitesse et, de ce fait, toujours inaccessibles à la compréhension. Cette recherche du mouvement atteint son apogée avec les séries photographiques d'Edward Muybridge, et celles d'Étienne-Jules Marey, qui décomposent la course d'un cheval, la marche d'un homme ou encore les acrobaties d'un gymnaste. L'humanité découvre alors, que dans sa course, le cheval ne décolle jamais ses quatre pattes en même temps, faisant mentir du même coup toute l'histoire de la peinture et affirmant de plus belle l'autorité de la reproduction photographique sur la représentation. Pourtant, si elle parvient à fixer le saut d'une danseuse mieux que n'a su le faire la peinture avant elle, la photographie se retrouve bloquée au même point : ses images demeurent désespérément fixes. André Malraux commente, dans *Le musée imaginaire*, les efforts de la peinture pour retranscrire le mouvement, en particulier ceux de la peinture baroque dont les figures tragiques sont le témoin d'une tentative acharnée pour ébranler la fixité de la toile. Ne pouvant user du temps, le baroque s'appuie sur la mise en scène de corps torturés, de visages grimaçants et de bouches hurlantes, pour hanter ses tableaux du spectre du mouvement. Bazin évoque cette tentative désespérée pour s'approprier la quatrième dimension tant désirée : « la perspective n'ayant résolu que le problème des formes non celui du mouvement, le réalisme devait se prolonger naturellement par une recherche de l'expression dramatique dans l'instant, sorte de quatrième dimension psychique capable de suggérer la vie dans l'immobilité torturée de l'art baroque »¹. Or, deux siècles plus tard, la photographie en est toujours réduite aux mêmes artifices que la peinture baroque. Si elle peut désormais décomposer les mouvements les plus rapides avec l'objectivité mécanique, le dynamisme

¹ *Ibid.*, p. 11.

qu'éprouve le spectateur à la vue des instantanés repose toujours sur une « quatrième dimension psychique ».

Pour trouver les racines du mouvement cinématographique, il nous faut nous éloigner momentanément de la photographie pour nous intéresser à l'animation. Dès le XVI^e siècle, voyagent en Europe des dispositifs comme la « lanterne magique », petits théâtres lumineux mettant en scène des silhouettes de papier découpé. Ces attractions très prisées s'appuient d'abord sur un mouvement manuel et continu, les opérateurs « animant » eux-mêmes, avec leurs mains, le mouvement des silhouettes à la manière d'un théâtre de marionnette. Mais peu à peu, les procédés s'affinent et se mécanisent. Les dispositifs aux noms baroques et mystérieux comme le phenakistiscope, le thaumatrope ou encore le praxinoscope, se multiplient et inventent l'« image en mouvement » — la succession rapide d'images fixes produisant l'illusion du mouvement continu. Toutes ces machines fonctionnent d'abord avec des dessins : c'est donc bien l'animation qui invente le cinéma, à l'époque où la photographie se débat encore avec l'immobilité de ses images. Les films d'animation constituent les premiers films, à l'instar de *Pauvre Pierrot*, réalisé en 1891 par Émile Reynaud, quatre ans avant l'invention du cinématographe des frères Lumière. Mais bien vite, la photographie prend sa revanche et le cinéma en « prise de vues réelles » se développe : d'abord le kinétographe, première caméra mise au point par Edison, et le kinétoscope, permettant le visionnage individuel des films du kinétographe ; et puis, le cinématographe des frères Lumière qui connaît un succès international, et invente le cinéma avec sa technique de projection publique. Après la photographie, le cinéma parachève la poursuite du réel en y ajoutant le temps : « Pour la première fois, l'image des choses est aussi celle de leur durée et comme la momie du changement. »¹. Cette nouvelle objectivité du cinéma en prise de vues réelles repose, non seulement sur la reproduction mécanique de la photographie, mais également, et c'est ce qui manquait jusque-là, sur la projection mécanique des images produites :

Ce n'est qu'au cours de la dernière décennie du XIX^e siècle que la génération automatique d'images et leur projection automatique ont finalement été combinées. Un œil mécanique s'est couplé à un cœur mécanique ; la photographie a rencontré le moteur. C'est ainsi qu'est né le cinéma, un régime très particulier du visible. L'irrégularité, la non-uniformité, l'accident et les autres traces du corps

¹ *Ibid.*, p. 14.

humain, qui accompagnaient auparavant inévitablement les projections d'images animées, ont été remplacés par l'uniformité de la vision mécanique.¹

Forte de cette objectivité redoublée, la prise de vues réelles s'impose rapidement par rapport à l'animation. Dans son livre *Le langage des nouveaux médias*, Lev Manovich montre à quel point le terme de « cinéma », dérivé du nom de l'appareil des frères Lumière, phagocyte le champ technique jusqu'à désigner l'image en mouvement en général. Sans doute le cinéma a-t-il besoin de se détacher de l'animation qui devient une de ses sous-catégories, de dissimuler ses liens avec ce cousin bâtard, pour embrasser pleinement son ontologie indicielle qui le place du côté du réel objectif, faisant de lui une prolongation, et même un double, de l'objet filmé. À la fin du XIXe, le cinéma constitue l'apogée de l'entreprise moderne de la figuration naturaliste. Le cinématographe est une machine à produire du réel, en témoigne cette fameuse anecdote : au cours de la première projection du film des frères Lumière *L'arrivée d'un train en gare de la Ciotat*, les spectateurs du premier rang se seraient levés dans un sursaut à la vue du train roulant sur eux. C'est que, comme l'écrit Deleuze, « le cinéma ne nous donne pas une image à laquelle il ajouterait du mouvement, il nous donne immédiatement une image-mouvement. »². Plus qu'un art de l'illusion porté à son paroxysme, le cinéma est l'art de l'immersion qu'il n'aura de cesse de poursuivre et d'augmenter tout au long du XXe siècle.

Le réalisme esthétique du cinéma : « de la reproduction à l'expression »

Si le terme de « cinéma » désigne, dès sa naissance, la technique de la prise de vues réelles, reléguant l'animation, qui avait pourtant inventé l'image en mouvement, à un rang subalterne, la séparation entre les deux techniques n'est pourtant pas aussi claire. Parmi les premiers films marquant septième art, nombre d'entre eux font un usage extensif des techniques d'animation, à l'instar des cinq cents et quelques courts métrages de Georges Méliès, considéré comme le père des effets spéciaux. Méliès redouble d'inventivité pour placer ses acteurs réels dans des situations et des décors plus fantastiques les uns que les autres, voyageant jusqu'à la Lune ou combattant des

¹ MANOVICH Lev, *The language of new media*, Cambridge, MIT Press, 2001, p. 251. Traduction personnelle « It was not until the last decade of the nineteenth century that the automatic generation of images and their automatic projection were finally combined. A mechanical eye became coupled with a mechanical heart; photography met the motor. As a result, cinema — a very particular regime of the visible — was born. Irregularity, non-uniformity, the accident and other traces of the human body, which previously inevitably accompanied moving image exhibitions, were replaced by the uniformity of machine vision. »

² DELEUZE Gilles, *Cinéma. 1, L'image-mouvement*, Paris, Éditions de Minuit, 2013. *Ibid.*, p. 11.

squelettes. Bien souvent, ces décors sont peints et les trucages reposent sur de l'animation. En parallèle de ce cinéma très plastique, les frères Lumière privilégient les vues naturelles et inaltérées, conformément à une logique naturaliste. À l'inverse des films de leurs principaux concurrents, Edison et Dickson qui tournent en studio, les frères Lumière envoient leurs opérateurs tourner dans des décors réels pour capturer des scènes de la vie courante. En résulte une série de films aux titres évocateurs : *La sortie de l'usine Lumière à Lyon*, *Les forgerons*, *La mer...* Bien vite, les opérateurs Lumière se mettent à voyager et rapportent des vues filmées dans les lointaines colonies françaises comme *Indonésie*, *Le Village de Namo : panorama pris d'une chaise à porteurs* tourné en Indochine. Le cinéma, tout comme la photographie avant lui, est employé ici comme témoignage du réel dans une approche parfois plus près de la science, ou de l'ethnographie, que de l'art. La conception du cinéma des frères Lumière diffère donc grandement de celle de Georges Méliès qui n'hésite pas à s'abstraire des limites physiques pour laisser libre court à sa créativité à grands renforts de trucages hérités des techniques d'animation. Pourtant, il y a un point sur lequel ces deux approches s'accordent : le cinéma doit toujours apparaître comme un enregistrement du réel. Manovich voit dans cette règle tacite la véritable séparation entre l'animation et le cinéma :

En revanche, le cinéma s'efforce d'effacer toute trace de son propre processus de production, y compris toute indication que les images que nous voyons pourraient avoir été construites plutôt qu'enregistrées. Il nie que la réalité qu'il montre n'existe souvent pas en dehors de l'image cinématographique, image obtenue en photographiant un espace déjà impossible, lui-même assemblé à l'aide de maquettes, de miroirs et de matte paintings, puis combiné à d'autres images par impression optique. Il prétend être un simple enregistrement d'une réalité déjà existante, tant pour le spectateur que pour lui-même.¹

La force du cinéma tient dans la réalité qu'il met — ou du moins qu'il prétend — mettre en scène. Ainsi, les effets spéciaux qui, à première vue, enfreignent grossièrement la règle s'efforcent pourtant de la respecter et d'en faire leur principal guide à travers les différentes expérimentations artistiques et techniques : il s'agit de faire oublier le lien avec l'animation et d'emporter la crédibilité du spectateur. L'image cinématographique doit toujours être l'empreinte d'un « ça a été ».

¹ L. Manovich, *The language of new media*, op. cit., p. 253. Traduction personnelle : « In contrast, cinema works hard to erase any traces of its own production process, including any indication that the images which we see could have been constructed rather than recorded. It denies that the reality it shows often does not exist outside of the film image, the image which was arrived at by photographing an already impossible space, itself put together with the use of models, mirrors, and matte paintings, and which was then combined with other images through optical printing. It pretends to be a simple recording of an already existing reality — both to a viewer and to itself. »

Pourtant, dans les premiers temps, cette conception indicielle du cinéma semble un frein à la créativité, non pas de la mise en scène très riche chez des réalisateurs comme Georges Méliès, mais celle des opérateurs caméra. C'est que les premiers *cameramen* comme « Edison, Lumière, Méliès, les opérateurs de Pathé ou Gaumont et les nombreux forains qui utilisent des caméras en France ou ailleurs conçoivent les films comme un « regard sur » les choses et non un « regard depuis » la conscience d'un observateur. »¹. Puisque le cinéma est compris comme un pur enregistrement du monde, quand bien même il mettrait en scène des mondes fantastiques à l'instar du *Voyage sur la Lune*, la simple notion de « réalisation » peut difficilement exister. Méliès a beau inventer des procédés qui préfigurent déjà le montage, comme l'arrêt de la caméra et le raccord-ellipse, il ne s'enfonce pas dans la scène pour varier ses cadrages, pas plus qu'un spectateur de théâtre ne monte sur les planches, en pleine représentation, pour examiner au plus près le visage tragique d'un acteur. À ce titre, on a employé l'expression de « théâtre filmé » à propos des premiers films dont la caméra frontale constitue un simple appareil d'enregistrement et non un outil au service de l'expression. C'est que le cinéma du début du XXe peine encore à s'affirmer comme un art. L'objectivité dont il est porteur semble parfois l'amener plus proche du compte-rendu – les « actualité cinématographiques » voient le jour en 1907 – ou de l'observation scientifique – à l'image des films ethnographiques des frères Lumière – comme la photographie avant lui. Peut-être cette vision est-elle due à ses dimensions mécanique et technologique qui, à l'époque, comme toujours aujourd'hui, sont facilement opposées à l'expression de la subjectivité. Quoi qu'il en soit, le problème de ce regard cinématographique proche de celui du spectateur de théâtre c'est qu'il « nous laisse devant l'image, sans guère nous inviter à entrer dedans. »². En un mot, ce qui manque aux premiers films c'est l'immersion, ou plutôt l'expression qui doit délivrer le cinéma de la reproduction photographique.

Pour Malraux, le libérateur du cinéma se nomme David W. Griffith. Celui-ci est considéré comme l'inventeur du montage, en particulier du gros plan, lors son film de 1915, *Naissance d'une nation*. Il est en réalité difficile d'attribuer l'entière paternité du découpage à Griffith tant les expérimentations en la matière ont commencé presque en même temps qu'étaient tournés les premiers films, vingt ans plus tôt ; mais les historiens s'accordent sur le rôle fondateur du réalisateur américain, tenu comme l'un des « pères du montage » aux côtés d'Eisenstein et de

¹ BARNIER Martin et JULLIER Laurent, *Une brève histoire du cinéma : 1895-2020*, S.I., Fayard/Pluriel, 2021, p. 31.

² *Ibid.*, p. 29.

Kouletchov. Ce qui importe vraiment, c'est la révolution que constitue le découpage et que Malraux résume en ces termes :

Tant que le cinéma n'était que le moyen de reproduction de personnages en mouvement, il n'était pas plus un art que la phonographie ou la photographie de reproduction. Dans un espace circonscrit, généralement une scène de théâtre véritable ou imaginaire, des acteurs évoluaient, représentaient une pièce ou une farce que l'appareil se bornait à enregistrer. La naissance du cinéma en tant que moyen d'expression (et non de reproduction) date de la destruction de cet *espace circonscrit*; de l'époque où le découpeur imagina la division de son récit en plans, envisagea, au lieu de photographier une pièce de théâtre, d'enregistrer une succession d'instantanés, d'approcher son appareil (donc de faire grandir les personnages dans l'écran quand c'était nécessaire), de le reculer; surtout de substituer au plateau d'un théâtre le «champ», l'espace qui sera limité par l'écran – le champ où l'acteur entre, d'où il sort, et que le metteur en scène *choisit*, au lieu d'en être prisonnier. Le moyen de reproduction du cinéma était la photo qui bougeait, mais son moyen d'expression, c'est la succession des plans.¹

Le découpage délivre le cinéma de ce regard figé et extérieur où l'avait enfermé la logique de la reproduction photographique. La caméra peut se déplacer, aller au plus près des événements, et s'incarner : elle est désormais le regard d'une conscience. Ce faisant, le cinéma cesse d'être un simple écran, une surface opaque sans épaisseur, pour devenir un espace dans lequel le spectateur peut s'immerger. Avec le montage alterné, il est désormais au plus près de l'action comme dans ces courses-poursuites qui font le succès du cinéma d'action ; avec le champ-contre-champ, il est entouré des personnages et il peut s'écrier comme le théoricien Béla Balázs : « je vois ce qu'ils voient, au lieu de voir depuis ma place »². C'est de cette liberté de mouvement, « de l'indépendance du cinéaste à l'égard de la scène de théâtre, que le cinéma [naît] en tant qu'art. »³. Vers 1920, le terme d'« opérateur de caméra » est devenu insuffisant, et le mot « réalisateur » entre enfin dans le lexique du cinéma. La décennie qui suit est une course en avant vers l'expressivité de ce nouveau langage cinématographique qui, bientôt, s'accompagne d'une grammaire. Le format du long-métrage, popularisé par *Naissance d'une nation*, se développe et les grosses productions hollywoodiennes se succèdent, à l'image des *Dix commandements* de Cecil B. Mill, ou de *Ben Hur* de Fred Niblo qui utilise pas moins de cinquante-deux caméras pour tourner la scène de la course char, preuve s'il en est que le montage joue désormais un rôle important. En Russie, des cinéastes comme Sergueï Eisenstein, Lev Koulechov ou encore Dziga Vertov poursuivent justement le travail

¹ MALRAUX André, « Esquisse d'une psychologie du cinéma », in *Cinématographe*, n° 24, 1977.

² BALÁZS Béla, *Béla Balázs: early film theory: Visible man and the spirit of film*, traduit par Rodney Livingstone, New York, Berghahn Books, 2011, p. 99.

³ MALRAUX André, *Le musée imaginaire*, Paris, Gallimard, « Collection folio Essais », 2006, 285 p, p. 75.

de Griffith en théorisant et en menant des expérimentations sur le découpage cinématographique, établissant une grammaire encore en vigueur aujourd'hui. En Europe, de nombreux artistes s'emparent de ce nouveau médium pour sans cesse en repousser les possibilités esthétiques et narratives. Buñuel et Dali, Fernand Léger, Man Ray, Abel Gance et tant d'autres n'hésitent pas à contester les règles à peine établies pour mettre au point de nouveaux procédés de mise-en-scène mais, surtout, pour produire des expériences inédites. Et la créativité artistique est toujours aidée en cela de la créativité technique, le cinéma mêlant l'une et l'autre de manière indissociable. En 1927 sort *Le chanteur de Jazz*, le premier film à intégrer des séquences sonores. Face au succès du parlant, les majors américaines se jettent sur cette nouvelle technologie et bientôt la majorité des films sont intégralement parlants. Si les premiers « talkies » font un emploi timide du son — les personnages parlant l'un après l'autre, la bande-son et les bruitages n'empiétant jamais sur les dialogues — très vite, la mise-en-scène s'approprie ce nouveau moyen d'expression : les personnages s'interrompent sans hésiter alors que la musique joue en arrière-plan. Peu à peu, la scène sonore du cinéma devient aussi riche que celle du monde pour le bonheur des spectateurs qui s'abîment un peu plus dans la profondeur de l'écran. Et l'amélioration technique du son est redoublée, une dizaine d'années plus tard, de l'émergence de la couleur dont *Autant en emporte le vent*, tourné dans les tons saturés du Technicolor, constitue l'un des premiers spécimens. À la veille de la Seconde Guerre mondiale, le cinéma a terminé sa mue, laissant derrière lui la reproduction photographique et son naturalisme, pour affirmer sa force d'expression qui lui permet d'atteindre un réalisme immersif sans précédent dans l'histoire de l'art. Mais, selon la formule consacrée « un grand pouvoir implique de grandes responsabilités », et, au moment où la guerre s'abat sur l'Europe avec son cortège d'horreurs, le cinéma est soumis à une question morale. Que faire de son réalisme face à une telle réalité ?

L'esthétique du réalisme : filmer la vie

Après trois films réalisés dans les studios mythiques de *Cinecittà* pendant la Seconde Guerre mondiale, Rossellini envisage une nouvelle forme de cinéma, poussé par une nécessité morale et spirituelle, mais également par la question plus prosaïque des moyens. *Cinecittà*, qui a servi d'entrepôt d'armes pour les allemands, est partiellement détruite avant de servir de camp pour les réfugiés : les cinéastes italiens sont livrés à eux-mêmes, obligés, s'ils veulent continuer à tourner, de quitter le cadre du studio. Rossellini en prend son parti et commence le tournage de *Rome, ville*

ouverte, bientôt érigé en manifeste du néoréalisme. Réalité crue de la guerre, décors réels, extérieurs, acteurs non-professionnels, large espace laissé à l'improvisation et à la sensibilité spontanée : le néoréalisme prend le contre-pied du *film-studio* dont les romances hollywoodiennes sont le pur produit, et se distingue par sa caractéristique fondamentale, la recherche de la vérité. Ce réalisme constitue une revendication artistique qui sonne comme un devoir moral face au cauchemar dont émerge à peine l'Europe. La réalité de Rossellini est celle d'une Italie à genoux, baignant dans la pauvreté, encore prise dans la torpeur où l'a plongée le fascisme vingt ans plus tôt ; c'est aussi celle de l'Allemagne rasée, de Berlin en ruines, et de la découverte des camps. Cette réalité apporte une contradiction sévère à l'humanisme et son idéal de progrès humain. « Écrire un poème après Auschwitz est barbare » écrit Adorno, et, si les néoréalistes n'abandonnent pas l'idée d'une certaine beauté, c'est en tout cas une beauté qui va chercher sa matière au plus de la misère et des laissés-pour-compte. Rossellini tourne d'abord *Païsa* et *Rome, ville ouverte* qui reconstituent la réalité du fascisme et de son opposition par les partisans puis *Allemagne, année zéro* qui présente une capitale allemande en ruines après la violence des bombardements alliés. Luchino Visconti et Vittorio de Sica, quant à eux, choisissent de montrer la réalité quotidienne de l'Italie d'après-guerre avec des films comme *La terre tremble* ou *Le voleur de bicyclette*. Le néoréalisme choque et impressionne par le réalisme cru qu'il déploie. Son style, mais surtout sa méthode de réalisation qui entretient des liens directs avec le documentaire, bien que les films néoréalistes demeurent des fictions qui ne se privent pas de ressorts mélodramatiques, marque une rupture par rapport au cinéma classique de l'âge d'or hollywoodien qui consacre des stars sublimées par une photographie soignée en studio. Avec le néoréalisme, le public découvre des acteurs et actrices souvent amateurs, des fins plus ambiguës que les *happy-end* du cinéma américain, des vies et des réalités bien proches de la sienne. Dans son article *Le réalisme cinématographique et l'école italienne de la libération*, André Bazin souligne la part importante qui est laissée à l'improvisation lors des tournages de Rossellini, de façon à toujours pouvoir saisir le réel au vol, inscrivant la démarche du néoréalisme à contre-courant des usages de la production cinématographique :

Depuis le parlant surtout, un film exige un travail trop complexe, engage trop d'argent pour admettre le moindre flottement en cours de route. On peut dire que le premier jour de tournage, le film est déjà virtuellement réalisé sur le découpage qui prévoit tout. Les conditions matérielles de la réalisation en Italie, immédiatement après la Libération, la nature des sujets traités et sans doute aussi quelque génie ethnique ont libéré les metteurs en scène de ces servitudes. Rossellini est parti

avec sa caméra, de la pellicule et des canevas de scénarios qu'il a modifié au gré de son inspiration, des moyens matériels ou humain, de la nature, des paysages...¹

Les films néoréalistes sont, en ce sens, des films *a posteriori*, réalisés d'après leur confrontation avec le réel, et non avant-même le tournage, dans le microcosme feutré d'un studio. Bazin relie cette méthode au réalisme des films russes des années 20 qui déjà recherchaient une forme de pureté du réel. Mais qu'on ne s'y trompe pas, le réalisme n'est pas un refus de l'esthétique, bien au contraire. C'est cette esthétique qui a avait fait le succès du réalisme soviétique d'Eisenstein, et c'est encore elle qui fait l'immense force du néoréalisme :

Ce ne sera pas, à mon sens, le moindre mérite du cinéma italien que d'avoir rappelé une fois de plus qu'il n'était pas de « réalisme » en art qui ne fût d'abord profondément « esthétique ». [...] Le réel comme l'imaginaire n'appartient, en art, qu'au seul artiste, la chair et le sang de la réalité ne sont pas plus aisés à retenir dans les filets de la littérature ou du cinéma que les fantaisies les plus gratuites de l'imagination. [...] [Le réalisme du cinéma italien] n'entraîne nullement une régression esthétique, mais au contraire un progrès de l'expression, une évolution conquérante du langage cinématographique, une extension de sa stylistique.²

Et, en effet, comment ne pas être pas touché par les moments de grâce qui parcourent les films du mouvement ? De la rencontre timide du GI Joe et de Carmela sous les étoiles de Sicile dans *Païsa* aux errances d'une Anna Magnani enceinte à travers la campagne italienne dans *L'amore*, sans oublier la fierté éclatante de Giuseppe et Pasquale qui défilent à cheval devant les autres enfants dans *Sciuscià*. Chacun des films de Rossellini, de De Sica ou de Visconti présentent, à un moment ou à un autre, une de ces fulgurances magnifiques, un instant d'éternité, qui tiennent dans cette esthétique faisant corps avec le réalisme d'une manière presque sensible, physique. Cette exaltation de la beauté dans le réel ne doit pas pour autant être prise pour un idéalisme qui prêterait à sourire ou pour une charmante naïveté. Le néoréalisme est né de la Seconde Guerre mondiale, et son horreur, loin d'être niée, sert de point de départ à une réflexion lucide sur l'humain. Lorsque dans *Païsa*, Rossellini filme longuement le cadavre d'un résistant flottant sur le Pô, et qu'il fait dire à un autre avant son exécution « *Je me pissurai dessus comme tous les autres* », il insiste sur la condition humaine — thème alors au centre de l'actualité intellectuelle, à l'image de l'absurde camusien ou

¹ A. Bazin, *Qu'est-ce que le cinéma ?*, op. cit., p. 274-275.

² *Ibid.*, p. 267-269.

de l'existentialisme — et la violence potentielle avec un réalisme cru qui ne fait pas de faveurs. Et pourtant, l'enterrement du résistant par ses camarades, l'espoir qui anime leur combat et débouche finalement sur la libération de l'Italie en dit assez sur le rôle fondamental de chaque individu lorsqu'il appartient à l'humanité. Cette humilité éclairée tranche avec le point de vue hollywoodien qui voudrait voir des héros sauver le monde. C'est ainsi que Rossellini et son scénariste, Federico Fellini, font le choix de décrire la reconquête de l'Italie, non sous le prisme de la grande histoire, mais sous celui des individualités humaines, le rôle infime des individus seuls, leurs rencontres, leurs vies, leurs morts. De ces morceaux d'existences anodins naissent des scènes magnifiques qui en deviennent universelles. Finalement chacun, par sa capacité à faire preuve d'humanité, contribue à redonner foi en l'humain, et le rôle des individualités se révèle soudain moins infime qu'il n'en avait l'air. Avec une telle morale, il n'est pas surprenant que le néoréalisme ait souvent été rapproché du christianisme, en particulier en vertu de sa dimension indiciaire, de sa capacité à révéler le réel. Proche de cette conception, et fortement inspiré par la phénoménologie et l'existentialisme de son temps, Bazin défend le « révélationisme » qu'il entend appliquer à l'ensemble du cinéma, mais qu'il tire en premier lieu du néoréalisme. Le critique distingue entre les réalisateurs qui « croient à la réalité » et ceux qui « croient à l'image », Rossellini et ses semblables appartenant à la première catégorie. Selon le révélationnisme, les images cinématographiques, en tant qu'empreintes, suscitent une forme d'époché qui permet d'accéder au réel, de le voir nu et débarrassé de tout le bagage de préconceptions qui en voile la vue dans l'expérience commune. On retrouve ici l'ontologie réaliste du cinéma défendue par Bazin : l'image photographique et, par extension, celle de la caméra se confond avec son objet, attribuant au cinéma une responsabilité morale : être l'art de la vérité. En ce sens, le révélationnisme s'apparente presque à une croyance religieuse dans le pouvoir de réalité des images. Les historiens Martin Barnier et Laurent Jullier, dans leur ouvrage *Une brève histoire du cinéma*, font le lien avec la conception chrétienne des images « achiropite », terme « tiré de l'adjectif acheiropoïètos (« non fait de main d'homme ») [...], néologisme forgé par saint Paul pour qualifier le corps du Christ ressuscité »¹. L'empreinte du cinéma serait du même ordre que celle du christ contre le Saint-Suaire, à laquelle Bazin fait référence. Il existe d'ailleurs une critique catholique du cinéma dont le père Amédée Aifre est l'un des représentants et qui rapproche le néoréalisme de la démarche catholique dans sa dimension révélatrice. Les images de ce cinéma nous révéleraient la réalité (catholique) du monde comme le

¹ M. Barnier et L. Jullier, *Une brève histoire du cinéma*, op. cit., p. 192.

corps du christ ressuscité, non fait de la main de l'homme, nous révèle le message chrétien. C'est que le néoréalisme parvient à nous faire ressentir le « mystère de l'être »¹.

Sans appartenir pleinement à cette critique catholique, Bazin, qui échangeait régulièrement avec le père Aifre, en partage l'idéal kérygmaticque qui lui fait distinguer entre le « véritable réalisme qui est besoin d'exprimer la signification à la fois concrète et essentielle du monde » et « le pseudo-réalisme du trompe l'œil (ou du trompe-l'esprit) qui se satisfait de l'illusion des formes. »². On comprend alors son manque d'amour pour le trucage et le dessin animé. L'esthétique du réalisme, chère au critique, apparaît toujours menacée par la dimension illusionniste du cinéma, présente depuis ses origines avec des figures comme celle de Méliès. Contre le cinéma-vérité, le cinéma-illusion, celui des trucages et des effets spéciaux, représente une part importante, sans doute majoritaire. Mais le réalisme est préservé par l'accord tacite qui lie ces deux cinémas pour que cet art qui les réunit soit toujours présenté comme l'enregistrement de ce qui a été, une pure prise de vues. Ainsi, les techniques d'illusion ont-elles toujours été dissimulées pour être confondues avec la prise de vues réelles et séparées en cela de l'animation. Mais, à l'ère du numérique et des *pixels* dont les effets spéciaux se sont emparés, le cinéma-illusion devient doublement trompeur. Si sa surface esthétique est plus convaincante que jamais, les créatures photoréalistes remplaçant les costumes de carton-pâte, il est pourtant difficile d'ignorer son origine virtuelle. Le soupçon pèse, plus que jamais, sur l'image cinématographique dont il est difficile de ne pas douter qu'elle est un « ça a été ». Cela signifie-t-il que le cinéma doit renoncer au réel ou que le numérique constitue pour lui une menace existentielle ? Cet art souvent décrit comme celui de « la vérité vingt-quatre fois par secondes » semble, en effet, basculer dans le mensonge avec les images de synthèse. Mais le tableau n'est pas si sombre. S'il lui est certes difficile de renoncer à l'objectivité naturaliste qui l'avait prévalu d'une forme d'autorité ontologique, le cinéma pourrait également saisir cette redéfinition numérique comme l'occasion d'une libération. La photographie a bien délivré la peinture de sa poursuite de la ressemblance, ouvrant la voie aux symbolismes du XXe siècle. De quoi le numérique pourrait-il délivrer le cinéma ? La réponse est sans doute à chercher du côté de l'ontologie numérique, ce fameux « virtuel » qu'on oppose si souvent au « réel ».

¹ ALBERA François, « Notes sur les rapports de l'Office catholique international du cinéma et la filmologie », in *Cinémas*, n° 2-3, vol. 19, 2009.

² A. Bazin, *Qu'est-ce que le cinéma ?*, op. cit., p. 11.

II. Réalité numérique

Le paradigme mécaniste

Le monde dans lequel nous vivons, celui qui a fait advenir l'informatique et qui abrite ses objets virtuels, est communément appelé « moderne ». Il s'agit bien d'un monde et pas simplement d'une époque. Nous l'avons déjà évoqué plus tôt, la Modernité désigne, plus qu'une phase temporelle, une manière d'être pour la société moderne qui lui fait voir, percevoir et finalement concevoir, au double sens passif et actif, un monde qui lui est propre. Cette idée d'une cosmologie particulière à chaque société correspond à ce qu'on a nommé le « tournant ontologique de l'anthropologie ». Contre l'idée ancienne d'une Nature uniforme constituant un fond commun sur lequel viendraient se juxtaposer des cultures différentes, cette approche anthropologique défend au contraire l'existence de couples « nature-culture », les deux termes étant indissociables. Chaque nature s'accompagne donc d'une culture propre et inversement, constituant ainsi différents types de « mondiations », de compositions de mondes qui ne sont pas de simples représentations mais des systèmes ontologiques:

De ce fait, une fois que le mouvement de mondiation est engagé pour un humain, c'est-à-dire dès la naissance, il ne produit pas une « vision du monde », c'est-à-dire une version parmi d'autres d'une réalité transcendante à laquelle seule la Science, ou Dieu, pourrait avoir un accès intégral ; il produit un monde au sens propre, saturé de sens et fourmillant de causalités multiples, qui chevauche sur ses marges d'autres mondes du même genre qui ont été actualisés par d'autres humains dans des circonstances analogues.¹

Philippe Descola défend l'idée que ce processus de mondiation passe par la ressemblance, ou la différence, que relèvent les humains entre eux-mêmes et les « non-humains » sur « un double plan, physique et moral »². L'anthropologue distingue ainsi quatre grands « modes d'identification » basés sur les continuités et discontinuités intérieures et extérieures : animisme, totémisme, analogisme et naturalisme.

Si la Renaissance constitue un moment charnière dans l'histoire de notre société c'est qu'elle est le point de bascule vers la Modernité. Nous l'avons vu, les artistes et scientifiques de

¹ P. Descola, *Les formes du visible*, op. cit., p. 11.

² *Ibid.*, p. 12.

l'époque rivalisent d'ingéniosité dans la poursuite de l'objectivation du monde à travers l'idéal déjà présent de la « figuration naturaliste ». Mais, malgré ces signes annonciateurs, au commencement de la Renaissance, le naturalisme est encore loin, ses prémisses dissimulées sous le fourmillement de l'analogisme qui bat alors son plein et qui repose, comme son nom l'indique, sur une discontinuité totale redoublée d'une analogie toute aussi totale : chaque entité du monde existe de manière purement individuelle et séparée des autres, reliée pourtant par des liens d'analogie qui poussent dans toutes les directions. Descola attribue cette ontologie si particulière à trois sociétés distinctes : la Renaissance, bien sûr, mais aussi la pensée chinoise et la pensée aztèque :

Malgré l'abîme culturel qui semblait séparer ces trois civilisations, elles étaient toutes trois obsédées par l'analogie comme un moyen de réduire le foisonnement des différences entre les objets du monde, leurs éléments constitutifs, les états, situations et qualités qui les décrivent et les propriétés dont on les dote en les connectant dans des réseaux élargis de correspondances. Dans les ontologies « analogistes » tout fait sens, tout renvoie à tout, aucune singularité ne demeure à l'écart des cheminements interprétatifs permettant de greffer une couleur sur une qualité morale, un jour de l'année sur une constellation ou un type d'humeur sur une fonction sociale.¹

Sans doute cette pensée analogique explique-t-elle le foisonnement des recherches artistiques, philosophiques et scientifiques de la Renaissance, leur entrecroisement constant et l'ingéniosité de leurs découvertes. Mais déjà, l'humanisme porte le sceau de la quatrième et dernière ontologie, celle du naturalisme. C'est à Descartes qu'on attribue généralement le coup de semonce de la Modernité avec son *Discours de la méthode* et sa célèbre incitation à connaître « la force et les actions du feu, de l'eau, de l'air, des astres, des cieux et de tous les autres corps qui nous environnent, aussi distinctement que nous connaissons les divers métiers de nos artisans »² afin de « nous rendre comme maîtres et possesseurs de la nature. »³ Soudain, Descartes rompt avec l'analogisme de la Renaissance en reconnaissant une continuité totale de la matière tout autant qu'il rompt avec l'animisme hérité de la pensée médiévale et antique en marquant une discontinuité fondamentale entre l'esprit humain et le reste du monde. Le monde cartésien se divise entre « substance pensante » — l'esprit humain, pure intériorité — et « substance étendue » — la matière, pure extériorité soumise aux lois physiques. Contre l'animisme qui sépare les corps et rapproche les

¹ *Ibid.*, p. 13.

² DESCARTES René, *Discours de la méthode*, Paris, Flammarion, « GF », 2000, 189 p, p. 85.

³ *Ibid.*

esprits, le naturalisme détache l'esprit humain, le plaçant hors de la nature, tout en rapprochant les corps:

C'est par leur esprit, non par leur corps, que les humains se distinguent des non-humains, notamment par cette intelligence réflexive de soi que l'on nomme cogito depuis Descartes; et c'est aussi par leur esprit, hypostasié en sorte d'âme collective, que les humains se distinguent les uns des autres dans des ensembles unifiés par le partage d'une langue, d'une culture, d'un système d'usages, d'une représentation du monde. Quant aux corps, vivants ou inertes, ils sont tous régis par des processus matériels, ils relèvent de l'universel à l'instar de l'espace, du temps et de la substance, en sorte que, malgré la taille disproportionnée du cerveau humain, ceux-ci n'ont pas vraiment une nature à part.¹

Si l'esprit humain est célébré pour son caractère si particulier, teinté de divin, jusqu'à donner forme à l'humanisme qui légitime les prérogatives de l'homme sur la nature, son corps, en revanche, n'est plus dans ce système qu'une simple instance de la « substance étendue ». On voit bien ici la bascule qui se joue avec la séparation de l'esprit et du corps. Le naturalisme vide le monde, le « désenchante » permettant du même coup « un développement sans précédent des sciences et des techniques »². Désormais, tous les phénomènes terrestres prennent une place dans des chaînes causales. C'est la naissance du mécanisme et, avec lui, la découverte des lois physiques et biologiques qui amènent une reconfiguration complète de notre manière de faire monde. La science moderne ne se satisfait plus des prétendus « miracles » qui expliquaient jusqu'à là bon nombre de phénomènes, elle part à la recherche des causes tangibles avec sa nouvelle méthode expérimentale. Galilée et Copernic remettent la Terre à sa place, Newton met à jour les lois de l'attraction et de la gravitation, Darwin montre que l'humain est le résultat d'une longue et lente évolution, battant en brèche la cosmogonie chrétienne. En quittant le royaume de Dieu, « la substance étendue » se libère de la loi divine pour n'être plus soumise qu'à la loi terrestre, celle de la physique. Même s'il prend soin de conserver la partie la plus noble de l'univers, l'esprit humain, sous le contrôle du divin, Descartes ouvre en réalité, et peut être sans le vouloir, la porte à un monde sans Dieu. C'est que, progressivement, la part matérielle du monde prend de plus en plus de place, empiétant sur sa part immatérielle. À partir du XVII^e siècle, tout phénomène de la nature devient explicable par une cause, elle-même effet d'une autre cause, etc... Dieu qui, jusque-là, était présent derrière chaque événement, se trouve soudain rejeté hors de la nature et réduit à la seule cause première. Dans ce

¹ P. Descola, *Les formes du visible*, op. cit., p. 434.

² *Ibid.*

monde mécaniste dont on s'émerveille de la perfection causale, et qu'on compare dès lors à une grande horloge bien huilée, Dieu est relégué à la place de « grand horloger ». Après avoir minutieusement conçu le mécanisme délicat de la nature, il ne lui reste plus qu'à « donner une chiquenaude pour mettre le monde en mouvement »¹, selon les mots de Pascal, et de se retirer à jamais.

Les mouvements qui rythment le monde moderne sont réglés selon des mécanismes auxquels n'échappent pas même les entités vivantes — excepté l'humain, mais pour combien de temps encore ? Descartes n'hésite pas à comparer les animaux à des automates qui perdent du même coup leur animalité au sens étymologique (*animal* provient du latin *anima* « souffle de la vie, principe vital »). Avec le causalisme cartésien, il n'y a plus d'impossibilité de principe à imaginer un animal d'un point de vue purement fonctionnel, amas d'organes et d'os semblables aux rouages et à l'armature d'une machine :

S'il y avait de telles machines qui eussent les organes et la figure d'un singe, ou de quelque autre animal sans raison, nous n'aurions aucun moyen pour reconnaître qu'elles ne seraient pas en tout de même nature que ces animaux.²

D'une manière plus générale, ce fantasme de l'automate propre au paradigme mécaniste trouve une manifestation concrète avec la révolution industrielle du XIX^e. Après avoir découpé le monde en entités discrètes agissant les unes sur les autres dans des processus physiques, il apparaît désormais possible de les manipuler et de les reconfigurer, presque à volonté, augmentant d'une manière inédite l'agentivité humaine. Jusqu'à quel point ? Après un premier stade de la machinisation au cours de la révolution industrielle, le philosophe Stéphane Vial décrit son second stade avec la révolution numérique :

Dans sa première forme historique, la machinisation est la mécanisation. Il s'agit de remplacer l'effort corporel et les opérations manuelles par des machines mécaniques, c'est-à-dire des appareils de métal, motorisés ou automatisés. Dans sa seconde forme, la machinisation est la numérisation. Il

¹ PASCAL Blaise, *Pensées*, Paris, Flammarion, « GF », 2015, p. 77. « Je ne puis pardonner à Descartes : il aurait bien voulu dans toute sa philosophie, se pouvoir passer de Dieu ; mais il n'a pu s'empêcher de lui faire donner une chiquenaude pour mettre le monde en mouvement ; après cela, il n'a plus que faire de Dieu. »

² R. Descartes, *Discours de la méthode*, op. cit., p. 92.

s'agit de remplacer l'effort intellectuel et cognitif par des machines numériques, qui traitent de l'information de manière massivement automatique.¹

En secondant les humains dans l'effort intellectuel, les ordinateurs font franchir à la machinisation un cap symbolique puisqu'elle touche au plus haut lieu de l'humanité, son cerveau, citadelle de son intellect et de sa conscience. En investiguant ainsi le champ cognitif, le numérique poursuit la désacralisation du monde entamée au début de l'ère moderne. Si Descartes avait démontré, par une expérience de pensée, qu'il n'y avait pas d'impossibilité logique à concevoir un animal comme un automate, il avait en revanche pris garde de ne pas franchir la ligne rouge de la conscience humaine, ce qui aurait équivalu à nier l'âme :

Car, on peut bien concevoir qu'une machine soit tellement bien faite qu'elle profère des paroles, et même qu'elle en profère quelques-unes à propos des actions corporelles qui causeront quelque changement en ses organes : comme, si on la touche en quelque endroit, qu'elle demande ce qu'on lui veut dire ; si en un autre, qu'elle crie qu'on lui fait mal, et choses semblables ; mais non pas qu'elle les arrange diversement, pour répondre au sens de tout ce qui se dira en sa présence, ainsi que les hommes les plus hébétés peuvent faire.²

Mais malgré les précautions de Descartes pour protéger l'homme de toute réduction mécaniste en le dotant d'un esprit, une brèche s'ouvre dans laquelle s'engouffre l'informatique.

Le paradigme numérique

L'ordinateur, dès ses débuts est conçu comme une sorte de cerveau artificiel, une machine permettant l'exécution automatique d'une grande variété d'opérations, d'abord purement mathématiques et puis, avec la naissance de l'informatique, plus largement intellectuelles, ne manipulant plus seulement des chiffres mais aussi et surtout de l'information. L'analogie avec le cerveau humain est particulièrement claire dans l'ouvrage *The computer and the brain*³, écrit en 1957 par John Von Neumann, mathématicien considéré comme l'un des pères fondateurs de

¹ VIAL Stéphane, *L'être et l'écran. Comment le numérique change la perception*, Paris, Presses Universitaires de France, 2013, 336 p, p. 68.

² R. Descartes, *Discours de la méthode*, op. cit., p. 92.

³ VON NEUMANN John, *The computer & the brain*, New Haven, Conn., Yale University Press, 2012.

l'informatique à qui l'on doit notamment l'architecture des ordinateurs que nous utilisons encore aujourd'hui. Dans son livre, Von Neumann s'appuie sur un argument fonctionnaliste pour dresser l'analogie : un neurone accomplit la même fonction qu'un transistor en transmettant une information binaire (allumé ou éteint). Ainsi Von Neumann écrit-il :

Le signal nerveux peut clairement être considéré comme un marqueur (à deux valeurs) dans le sens évoqué précédemment : l'absence de signal représente alors une valeur (par exemple, le chiffre binaire 0), et la présence d'un signal représente l'autre valeur (par exemple, le chiffre binaire 1)¹.

Ce qui lui permet de conclure : « les neurones apparaissent [...] comme l'organe logique de base - et donc aussi comme les organes numériques de base »². À travers ces quelques lignes, Von Neumann exprime clairement le paradigme fonctionnaliste à la base de l'informatique en s'attaquant au symbole le plus fort : le cerveau humain, siège de sa conscience. Il s'agit ici de réduire le réel à un ensemble de fonctions permettant d'obtenir des « effets-images » à partir de « causes-antécédents ». L'ordinateur de Von Neumann reprend ainsi à son compte l'ambition ancienne de l'automatisation de l'effort intellectuel et cognitif qui caractérise l'histoire de l'informatique.

Dans son livre *Computing: A Concise History*³, l'historien de la technologie Paul Ceruzzi décrit la naissance de l'ordinateur comme le résultat de la convergence de différentes automatisations : celle du calcul, de l'exécution d'une séquence d'opérations, du stockage et de la récupération d'information sous forme codée. En ce qui concerne l'automatisation du calcul, elle est très ancienne et prend ses racines bien avant l'avènement de la modernité. Howard Aiken, informaticien américain, offre un bon aperçu de sa riche histoire dans son article *Proposed automatic calculating machine*⁴, présenté à IBM en 1937:

Le désir d'accélérer le calcul mental et d'en réduire l'effort intellectuel, ainsi que celui d'éliminer les erreurs dues à la faillibilité de l'esprit humain, est probablement aussi vieux que l'histoire des mathématiques. Ce désir engendra la création et la construction d'abaques de toutes sortes,

¹ *Ibid.*, p. 43. Traduction personnelle : « The nervous pulse can clearly be viewed as (two-valued) markers in the sense discussed previously: the absence of a pulse then represents one value (say, the binary digit 0), and the presence of one represents the other (say, the binary digit 1). »

² *Ibid.*, p. 54. Traduction personnelle : « the neurons appear [...] as the basic logical organ – and hence also as the basic digital organs »

³ CERUZZI Paul E., *Computing: a concise history*, Cambridge (Mass.), MIT press, « The MIT press essential knowledge series », 2012.

⁴ AIKEN Howard Hathaway, « A Proposed Automatic Calculating Machine (1937) » in *Ideas That Created the Future: Classic Papers of Computer Science*, Harry R. Lewis (éd.), s.l., The MIT Press, 2021.

commençant avec le groupement de petits objets, comme des cailloux, utilisés initialement en petits tas, puis utilisés comme compteurs sur des planches compartimentées et enfin couissant sur des fils montés dans un cadre comme dans le boulier. Cet instrument fut vraisemblablement inventé par les Sémites, puis adopté en Inde, d'où il se propagea à l'Ouest vers l'Europe et à l'Est vers la Chine et le Japon. Après le développement du boulier, il n'y eut aucune amélioration des abaqes jusqu'à l'invention par John Napier de ses bâtons, ou réglettes de Neper, en 1617. Ces réglettes furent incorporées dans des instruments de formes diverses, parfois approchant le début de la mécanisation du calcul, mais c'est Blaise Pascal qui nous donnera la première machine à calculer au sens du terme que nous utilisons aujourd'hui.

La *Pascaline* mise au point dans les années 1640 marque une rupture avec les nombreux outils pour le calcul qui avaient été conçus jusque-là. La machine de Pascal est considérée comme la première machine à calculer, ouvrant la voie au calcul mécanique. De nombreuses machines suivent, comme celle de Leibniz qui ajoute la fonction de la multiplication à celle de l'addition, mais c'est véritablement à l'époque de la révolution industrielle que les machines à calculer prennent leur essor et commencent à être produites à grande échelle. Au même moment, une nouvelle étape importante est franchie en direction de l'ordinateur avec la machine analytique du mathématicien anglais Charles Babbage. Sa conception dans les années 1830 marque un accomplissement majeur dans l'automatisation d'une séquence d'opérations. Réutilisant les cartes perforées des métiers à tisser Jacquard mis au point quelques années plus tôt, Babbage conçoit ainsi la première machine à calculer programmable. Ada Lovelace, fille du poète Byron, écrit en 1843 un programme destiné à la « machine analytique » et devient ainsi la première programmeuse informatique de l'histoire. Les cartes perforées des métiers Jacquard continuent d'inspirer les inventeurs du XIXe et contribuent à la mise au point de la troisième automatisation nécessaire à l'avènement de l'ordinateur : le stockage et la récupération d'information sous forme codée. L'ingénieur Herman Hollerith développe un système de stockage de l'information utilisant les cartes perforées pour le recensement américain de 1890. C'est que les cartes perforées stockent l'information de manière numérique et présentent un avantage important : elles permettent à la fois le stockage, le contrôle et le calcul de l'information. Hollerith commercialise ensuite son invention en créant une entreprise qui deviendra, à force de reconfigurations, *IBM*. Ces différentes automatisations donnent lieu à l'invention de nombreuses machines utilisées à grande échelle pour calculer, communiquer ou enregistrer de l'information. Mais à l'aube des années 1930, aucune machine n'a encore fait la jonction entre ces différents types d'automatisation.

C'est dans la décennie qui suit que l'ordinateur fait une apparition presque simultanée avec les travaux de Konrad Zuse et d'Alan Turing, dans une approche pratique pour le premier, théorique

pour le second. Konrad Zuse, ingénieur allemand, se heurte comme nombre de ses collègues aux limites du calcul mécanique qui impose un découpage précis des opérations et une certaine lenteur d'exécution. Peu à peu, Zuse commence à formuler l'idée d'une machine universelle qui pourrait, à la manière d'un cerveau, exécuter une variété potentiellement infinie d'opérations. Paul Ceruzzi cite un extrait du journal de l'ingénieur en 1937 :

Depuis environ un an, je réfléchis au concept d'un cerveau mécanique... La découverte qu'il existe des opérations élémentaires permettant de résoudre tous les processus arithmétiques et cognitifs... Pour chaque problème à résoudre, il doit exister un cerveau spécialisé qui le résout aussi rapidement que possible.¹

Zuse comprend que les opérations de calcul, de stockage, de contrôle et de transmission de l'information, qui étaient jusqu'à là menées de manière séparée, étaient en fait équivalentes:

Zuse s'est rendu compte qu'il pouvait concevoir des dispositifs mécaniques pouvant être réorganisés de manière flexible afin de résoudre une grande variété de problèmes : certains nécessitant davantage de calculs, d'autres davantage de stockage, mais tous nécessitant différents degrés de contrôle automatique. En bref, il a imaginé une machine universelle.²

Cependant, l'expression de « machine universelle » si importante dans l'histoire de l'informatique en ce qu'elle a constitué la première description d'un ordinateur n'est pas de Konrad Zuse mais du mathématicien anglais Alan Turing. Dans son article de 1936, *On computable numbers, with an application to the Entscheidungsproblem*³, Turing résout un vieux problème mathématique en faisant l'expérience de pensée d'une « machine universelle » pouvant être reconfigurée en permanence de manière à accomplir n'importe quel calcul. C'est le principe même de l'ordinateur :

La définition classique d'une machine est celle d'un appareil qui fait une seule chose, mais la fait bien ; un ordinateur est en revanche une machine universelle, dont les applications ne cessent de surprendre ses créateurs. Turing a formalisé ce que Zuse avait reconnu d'un point de vue technique :

¹ BAUER Friedrich L., *The Computer - My Life*, Berlin, Heidelberg, Springer Berlin / Heidelberg, 1993, p. 44., Traduction personnelle : « For about a year now I have been considering the concept of a mechanical brain. . . . Discovery that there are elementary operations for which all arithmetic and thought processes can be solved. . . For every problem to be solved there must be a special purpose brain that solves it as fast as possible. »

² P.E. Ceruzzi, *Computing, op. cit.*, p. 24., Traduction personnelle : « Zuse realized that he could design mechanical devices that could be flexibly rearranged to solve a wide variety of problems: some requiring more calculation and others requiring more storage, though each requiring varying degrees of automatic control. In short, he conceived of a universal machine. ».

³ TURING A. M., « On Computable Numbers, with an Application to the Entscheidungsproblem », in *Proceedings of the London Mathematical Society*, n° 1, vol. s2-42, 1937, p. 230-265.

un ordinateur polyvalent, lorsqu'il est équipé d'un programme adapté, devient, selon les termes de Zuse, « un cerveau spécialisé » qui fait une seule chose, à savoir ce que le programmeur veut qu'il fasse.¹

C'est donc avec le concept de « machine universelle » que naît véritablement l'ordinateur. Mais il reste encore à le mettre en pratique : comment une seule machine pourrait-elle parvenir à accomplir une variété virtuellement infinie d'opérations ?

La réponse au problème que pose la « machine universelle » se trouve dans l'adoption du système binaire, qui compte en base 2 et ne requiert ainsi que l'utilisation de deux symboles : le 0 et le 1. Malgré son apparente simplicité, le système binaire permet l'émergence du sens et la description d'une quantité d'information potentiellement infinie. En effet, un couple binaire constitue l'entité logique minimale : si un système « unaire » n'émettant qu'un seul message, toujours le même quelles que soient les circonstances, ne peut produire aucune information, c'est à partir de l'introduction d'un deuxième message, d'une variance, que l'information surgit. Un individu qui ne saurait dire que « oui » n'apprendrait rien à ses interlocuteurs, mais dès lors qu'au « oui » s'ajouterait la possibilité du « non », alors il serait en mesure de communiquer de manière, certes plus laborieuse que le langage naturel, mais presque aussi efficace. Paul Ceruzzi parle de « paradigme numérique » pour décrire notre époque informatique marquée par le passage, pour les machines, du système décimal, utilisé depuis 2000 ans, au système binaire. Tout comme l'automatisation du calcul, on peut retracer les origines du comptage en base 2 bien en amont de la Modernité, notamment en Chine, avant sa réactualisation, plus proche de nous, par le mathématicien britannique George Boole et ses travaux sur la logique binaire, et puis surtout par Leibniz souvent crédité comme l'inventeur du système binaire moderne. Avec sa simplicité logique réduite à l'essentiel, le binaire s'accorde particulièrement bien avec le fonctionnement des machines — il est plus simple de raccorder les deux valeurs du binaire à la position « levé » ou « baissé » d'un interrupteur, que de mettre au point un appareillage de boutons pour rendre compte des dix positions requises par le système décimal. Surtout, comme le décrit le philosophe Luciano Floridi

¹ P.E. Ceruzzi, *Computing, op. cit.*, p. 27., Traduction personnelle : « The classical definition of machine is of a device that does one thing and one thing well; a computer is by contrast a universal machine, whose applications continue to surprise its creators. Turing formalized what Zuse had recognized from an engineer's point of view: a general-purpose computer, when loaded with a suitable program, becomes "a special purpose brain" in Zuse's words that does one thing —whatever the programmer wants it to do. »

dans son ouvrage *Information : A Very Short Introduction*¹, les ordinateurs parviennent parfaitement à discriminer entre deux états :

La seule trace d'intelligence que tout le monde s'accorde à attribuer à un ordinateur concerne sans controverse la capacité de ses dispositifs et circuits à distinguer les données binaires. Si l'on peut dire qu'un ordinateur perçoit quoi que ce soit, c'est la différence entre une tension élevée et une tension faible, en fonction de laquelle ses circuits sont programmés pour fonctionner.²

Les premiers ordinateurs numériques, fonctionnant sur le système binaire, sont développés au début des années 1940. Le terme de « digital » est utilisé pour la première fois par le mathématicien George Stibitz en 1942. Dans le cadre de la Seconde Guerre mondiale, la recherche en informatique bondit aux États-Unis sous l'impulsion des programmes gouvernementaux. C'est notamment à cette époque que sont inventés les transistors, permettant de remplacer les systèmes mécaniques des anciens calculateurs, trop lents pour la rapidité d'action requise par l'urgence de la guerre. Peu à peu, l'ordinateur numérique s'impose et, si les ordinateurs analogiques continuent toujours d'être utilisés pour leur grande fiabilité, leur mode de fonctionnement monotâche les condamne à une relégation certaine. L'ordinateur, conçu comme « machine universelle », ne peut être que l'ordinateur numérique.

Après la guerre, la « computer science » se développe rapidement. En langue française, l'ingénieur Philippe Dreyfus propose le terme d'« informatique » en 1962. Après plusieurs années de dissensus quant à la nature de cette nouvelle discipline, l'académie française tranche en apportant sa définition : « Science du traitement rationnel, notamment par machines automatiques, de l'information considérée comme le support des connaissances humaines et des communications dans les domaines techniques, économique et social ». Cette notion de « traitement de l'information » ouvre le champ des calculateurs désormais rebaptisés en « ordinateurs ». Il devient clair que l'informatique a une vocation bien plus large que les opérations de comptabilité. Dans le sillage de la cybernétique développée dans les années 1940 sous l'égide, notamment, de Norbert

¹ FLORIDI Luciano, *Information: a very short introduction*, Oxford, Oxford University press, « Very short introductions », 2010.

² *Ibid.*, p. 43., Traduction personnelle : « The only glimpse of intelligence everyone is ready to attribute to a computer uncontroversially concerns the capacity of its devices and circuits to discriminate between binary data. If a computer can be said to perceive anything at all, this is the difference between a high and a low voltage according to which its circuits are then programmed to behave. »

Wiener et son livre *Cybernetics: or Control and communication in the animal and the machine*¹, l'information prend une importance nouvelle :

Wiener est l'un des premiers, sinon le premier à mettre en évidence la notion d'information et à donner à cette notion un caractère très large et une portée quasiment universelle, aux côtés d'autres grandes notions comme l'énergie ou la matière.²

L'information devient une entité de base pour comprendre le monde. Luciano Floridi définit l'information d'une manière générale comme le couplage de la donnée, ou « data », et du sens. Si le « symbol grounding problem », c'est-à-dire la question de l'ancrage du sens dans la *data*, est une question ardue n'offrant aucune résolution certaine, la *data* elle-même peut être interprétée d'au moins trois manières :

Premièrement, la *data* peut être un manque d'uniformité dans le monde réel, [...] par exemple une lumière rouge sur un fond sombre [...] Deuxièmement, la *data* peut être un manque d'uniformité entre (la perception d') au moins deux états physiques d'un système ou de signaux. Par exemple, une charge plus ou moins élevée dans une batterie, un signal électrique variable dans une conversation téléphonique, ou le point et la ligne dans l'alphabet Morse. [...] Enfin, la *data* peut être un manque d'uniformité entre deux symboles, par exemple les lettres B et P dans l'alphabet latin.³

L'information, est donc avant tout un jeu de différences, à l'image de la langue qui en constitue un exemple paradigmatique.

Vers la fin des années 1940, le mathématicien Claude Shannon s'empare de cette notion d'information pour fonder sa « théorie mathématique de l'information » dans un article du même nom. Cette théorie ne s'intéresse pas tant au contenu de l'information qu'à sa transmission en elle-même, posant la double question de la quantité d'information transmise et de l'efficacité de son encodage. C'est à l'occasion de cet article que le concept de « bit » prend une résonance nouvelle. Un *bit* désigne la quantité minimale d'information pouvant être transmise. Ainsi, plus un message

¹ WIENER Norbert, *Cybernetics: or; Control and communication in the animal and the machine*, Second edition, 2019 reissue., Cambridge, Massachusetts, The MIT Press, 2019, 303 p.

² Philippe Breton, « L'informatique comme discipline existe-t-elle? Histoire d'un clivage qui sépare les informaticiens », *Réseaux* 24, no 3 (1987): 65-75.p.74

³ L. Floridi, *Information, op. cit.*, p. 38., Traduction personnelle : « First, data can be lacks of uniformity in the real world, [...] e.g. a red light against a dark background[...] Second, data can be lacks of uniformity between (the perception of) at least two physical states of a system or signals. Examples include a higher or lower charge in a battery, a variable electrical signal in a telephone conversation, or the dot and the line in the Morse alphabet.[...] Finally, data can be lacks of uniformity between two symbols, for example the letters B and P in the Latin alphabet. »

comporte de *bits*, plus la quantité d'information augmente et plus l'« uncertainty », i. e. l'information manquante, diminue. La théorie mathématique de l'information s'appuie donc sur cette unité minimale qu'est le *bit* pour quantifier l'information transmise par un système et étudier la manière dont elle peut être encodée. En guise d'exemple, une pièce qu'on tire à pile ou face produit un *bit* d'information dans un cas idéal — lorsque la pièce est parfaitement lisse, équilibrée et que sa probabilité de tomber sur l'une ou l'autre de ses faces est équiprobable. Dans le monde physique, il n'existe pas de telle pièce, et la quantité d'information transmise est toujours plus ou moins inférieure au *bit*, d'autant plus proche de zéro que la pièce est biaisée (en effet, une pièce qui tomberait toujours du même côté ne serait plus un système binaire, mais unaire ne transmettant dès lors aucune information). Et de la même manière, la quantité d'information qu'un ordinateur peut transmettre augmente avec le nombre de ses transistors, chacun d'entre eux étant en mesure de transmettre un *bit* d'information. Floridi propose une mesure assez simple pour quantifier la quantité d'information transmise par un système en général. Il suffit de compter le nombre de « questions oui/non » (« yes-no questions ») nécessaires pour parvenir à l'information demandée :

Dans la théorie mathématique de la communication (MTC), l'information n'est qu'une sélection d'un symbole parmi un ensemble de symboles possibles. Ainsi, pour comprendre simplement comment la MTC quantifie l'information, il suffit de considérer le nombre de questions oui/non nécessaires pour déterminer ce que la source communique. Une seule question suffit pour déterminer le résultat d'un lancer de pièce équitable, qui produit donc un *bit* d'information.¹

Le miracle de l'informatique, c'est finalement l'encodage d'une quantité d'information potentiellement infinie et d'apparence continue d'après cette unité minimale et discrète qu'est le *bit*. À partir de huit *bits*, i. e. une série de huit 0 et 1, la quantité d'information transmise est d'un *byte* permettant la génération d'une table de 256 possibilités. En combinant, les séquences de *bits*, il devient possible de décrire virtuellement tous les phénomènes existants et de donner une approximation, infiniment précise, de la continuité. Plus rien n'empêche alors l'informatique d'investir le « monde mouvant des phénomènes » en le décomposant et en mettant à jour l'information qui le sous-tend. Celle-ci reprend à son compte la conception scientifique du réel advenue avec la Modernité : un réel discret et donc manipulable. Mais dans cette science nouvelle il

¹ *Ibid.*, p. 59-60., Traduction personnelle : « For the mathematical theory of communication (MTC), information is only a selection of one symbol from a set of possible symbols, so a simple way of grasping how MTC quantifies information is by considering the number of yes/no questions required to determine what the source is communicating. One question is sufficient to determine the output of a fair coin, which therefore is said to produce one bit of information. »

n'est plus question de matière ou d'énergie : c'est l'information qui recouvre tout avec son nouvel atome numérique, le *bit*. Forte de cette nouvelle approche, l'informatique investit avec un grand succès le monde phénoménal dont elle livre une reproduction de plus en plus crédible, semblable en cela au cinéma dont la continuité discrète avait ébloui Bergson au début du XXe siècle :

C'est parce que la bande cinématographique se déroule, amenant, tour à tour, les diverses photographies de la scène à se continuer les unes les autres, que chaque acteur de cette scène reconquiert sa mobilité : il enfle toutes ses attitudes successives sur l'invisible mouvement de la bande cinématographique. [...] Tel est l'artifice du cinématographe.¹

Informatique graphique : des images aux pouvoirs nouveaux

Les « super-ordinateurs » des années 1950, grands comme des pièces entières, sont d'abord employé à faire des comptes : l'ordinateur, à son commencement, ne trahit pas ses origines, c'est une pure machine à calculer. Mais bien vite, les calculs s'affinent et s'effacent, ils s'enfoncent peu à peu sous une surface qui les dissimule à la vue du grand public. Le calcul informatique ne se laisse plus percevoir en tant que calcul. Les ordinateurs se dotent d'écrans qui affichent des images pixelisées en deux dimensions, offertes à l'utilisateur qui peut interagir avec elles. Les concepteurs informatiques, à l'instar de Steve Jobs, saisissent l'importance de la dimension phénoménale des ordinateurs. Les premiers *macintoshs* d'*Apple* se démarquent de la concurrence en offrant une interface graphique au lieu d'une interface en lignes de commande, avec des icônes intuitives comme la « corbeille » ou le « finder » en forme de chemise cartonnée, et des instruments d'interaction comme la « souris ». L'émergence et l'immense succès de ces interfaces signe l'entrée de l'informatique dans le champ de l'esthétique. Les *designers UX* (*User Experience*) et *UI* (*User Interface*) investissent ce nouvel espace et redéfinissent le design. Il s'agit désormais de *designer*, non plus des objets physiques, mais des objets virtuels et l'expérience qu'ils procurent. À cette fin, il importe de bien connaître les propriétés de ces nouveaux objets. Lev Manovich dans *Le langage des nouveaux médias* en distingue cinq fondamentales : la discontinuité, la modularité, l'automatisation, la variabilité et le transcodage.

Nous avons déjà évoqué la discontinuité à l'occasion de la description du système binaire. Le fonctionnement des ordinateurs numériques repose principalement sur leur capacité à

¹ BERGSON Henri, *L'évolution créatrice*, Réimpression de la 12^e éd., Paris, Presses universitaires de France, « Quadrige », 2013, p. 304-305.

discriminer entre deux états différents — 0 et 1 au niveau logiciel, « allumé » et « éteint » au niveau électronique. Les images numériques reposent donc sur ce fonctionnement discontinu qui les distingue des images argentiques capturées sur la pellicule. Avec les *raster images* (de l'anglais *raster* qui veut dire « trame ») naît le concept de « pixel » qui divise chaque image en un quadrillage de *picture elements*. Tout l'enjeu de la rastérisation consiste alors à traduire en discontinuité la continuité : la nouvelle image numérique est décrite sous la forme d'une séquence de *bits* codant la suite de *pixels* qui la compose. Chaque *picture element* comporte trois valeurs numériques pour les trois couleurs primaires, à partir desquelles l'ensemble des couleurs peut être reconstitué. Par exemple, une image au format *Bitmap 24 bits* attribue 8 *bits* à chaque canal RGB, ce qui permet d'obtenir 256 nuances de chaque couleur primaire, et au total, sur l'image, 16 777 216 couleurs différentes. Avec un tel nombre, l'illusion de continuité est saisissante : une image numérique dans un ce format ne procure pas, à celui qui la visionne, le sentiment d'une réduction du réel mais plutôt d'une transcription fidèle.

Puisque l'image pixelisée existe sous forme d'une suite de 0 et de 1 stockés dans la mémoire de l'ordinateur, elle est manipulable à souhait, ce qui nous amène à la deuxième caractéristique des objets numériques évoquée par Manovich : leur modularité. Du fait de sa discontinuité, chacun des éléments qui composent une image numérique — c'est-à-dire chaque *pixels* et même, plus profondément, chaque *bit* — peut être extrait, modifié, supprimé, déplacé ou remplacé. C'est le propre de l'informatique graphique que de favoriser la manipulation au sein même des images. On comprend alors son adoption par l'industrie des effets spéciaux. D'une manière générale, les objets numériques sont composés de différentes entités, de la plus grande à la plus petite, qui gardent leur autonomie :

Les éléments multimédias, qu'il s'agisse d'images, de sons, de formes ou de comportements, sont représentés sous forme d'ensembles d'échantillons discrets (pixels, polygones, voxels, caractères, scripts). Ces éléments sont assemblés en objets de plus grande taille, mais ils conservent leur identité distincte. Les objets eux-mêmes peuvent être combinés pour former des objets encore plus grands, sans pour autant perdre leur indépendance.¹

¹ L. Manovich, *The language of new media*, op. cit., p. 51., Traduction personnelle : « Media elements, be it images, sounds, shapes, or behaviors, are represented as collections of discrete samples (pixels, polygons, voxels, characters, scripts). These elements are assembled into larger-scale objects but they continue to maintain their separate identity. The objects themselves can be combined into even larger objects — again, without losing their independence. »

Les deux premières propriétés des objets numériques, la discontinuité et la modularité, amènent de manière logique à la troisième : l'automatisation. Puisque ces objets virtuels sont décomposables en entités de plus en plus infimes qu'on peut moduler à volonté, il devient possible d'automatiser leur création. On retrouve ici, l'idéal photographique de l'automatisme de la prise de vues comme garantie contre l'immixtion d'une subjectivité humaine. Mais dans le cas du numérique, la situation est plus ambiguë. Si le processus de rendu d'une image 3D se déroule certes de manière « mécanique », dans une sorte de « boîte noire » qui ne permet pas d'intervention humain au cours du processus, il n'en demeure pas moins que l'algorithme de rendu a été codé par des développeurs humains et que le monde virtuel a été assemblé par des infographistes.

La quatrième caractéristique des objets numériques, la variabilité, pousse un cran plus loin le phénomène observé par Walter Benjamin dans son célèbre texte *L'œuvre d'art à l'époque de sa reproductibilité technique*. Si Benjamin dresse le constat qu'à l'ère de la mécanisation les médias peuvent être reproduits de manière quasi-infinie à l'identique, et leurs millions de copies distribuées de par le monde, permettant à des millions de spectateurs de voir la même œuvre à l'instar du cinéma ; les « nouveaux médias » que décrit Manovich introduisent la possibilité de la variation dans la copie. Il devient possible de distribuer des œuvres qui varient en taille, en format, en composantes. Et même plus, la variation dynamique fait partie des objets virtuels eux-mêmes. Ainsi, un jeu vidéo 3D rendu en temps réel ne présente-t-il jamais deux fois exactement la même séquence d'images, à l'inverse d'un film. Puisque le jeu doit pouvoir répondre aux actions du joueurs, son contenu ne peut pas préexister à la partie mais doit être généré à chaque instant. Soyons clair, il ne s'agit pas de dire que le jeu n'existe sous aucune forme avant qu'il ne soit « lancé » : en réalité, le joueur achète un programme qui contient, dans son code, une marge de variation plus ou moins grande selon la complexité du jeu. C'est cette variation qui est calculée en temps réel, procurant l'illusion de « mondes ouverts ». D'une manière générale, l'informatique cherche à représenter la *data*, et le monde lui-même, « en tant que variables plutôt que constantes »¹.

Enfin, la dernière propriété du « transcodage » tient au fait que, sur un support informatique, tous les médias se changent en données numériques. La nature des objets virtuels est donc double : d'une part, leur nature phénoménale, au niveau de la perception humaine, et d'autre part, leur existence numérique, suite de 0 et de 1 stockée dans un fichier informatique. Dans la « cosmogonie » de l'ordinateur, une image informatique n'est donc rien d'autre qu'une succession de *bits*. Il y a une tension, presque une opposition entre l'image perçue et ses dessous :

¹ *Ibid.*, p. 61.

La structure d'une image informatique en est un bon exemple. Au niveau de la représentation, elle appartient au domaine de la culture humaine, entrant automatiquement en dialogue avec d'autres images, d'autres « semes » et « mythemes » culturels. Mais à un autre niveau, il s'agit d'un fichier informatique composé d'un en-tête lisible par une machine, suivi de chiffres représentant les valeurs RGB de ses pixels. À ce niveau, elle entre en dialogue avec d'autres fichiers informatiques. Les dimensions de ce dialogue ne sont pas le contenu, les significations ou les qualités formelles de l'image, mais la taille du fichier, le type de fichier, le type de compression utilisé, le format du fichier, etc. En bref, ces dimensions relèvent davantage de la cosmogonie propre à l'ordinateur que de la culture humaine.¹

Manovich distingue alors entre une « couche culturelle » et une « couche informatique », constituant l'« ontologie de l'ordinateur », chacune d'elles s'influençant réciproquement. Si l'on peut comparer nouveaux et anciens médias sur la base de leur apparence, on ne peut pas pour autant ignorer cette couche profonde, ces fondations numériques qui soutiennent la programmabilité des objets virtuels. C'est que cette dimension programmable finit par remonter à la surface et affecter la phénoménalité des nouveaux médias : l'expérience d'un jeu vidéo, par exemple, tient autant à son apparence visuelle qu'à son interactivité. En évoquant la propriété de « transcodage » des nouveaux médias et leur double nature, culturelle et informatique, Manovich réactive une très vieille opposition de la philosophie : celle de la surface et de la profondeur selon laquelle la réalité se cacherait sous un voile trompeur. Et cette idée est encore accentuée par le fait que l'image numérique devient une interface : l'écran agit comme une fenêtre ouvrant sur une « réalité virtuelle », objet de tous les fantasmes.

Le réel du virtuel

Le terme de « virtuel » sert souvent de synonyme pour désigner le numérique. Mais cette utilisation pose problème pour son manque de clarté et la confusion qu'elle entraîne. Dans une première appréciation, largement répandue, le virtuel sert à désigner le faux, l'apparence mensongère dont le numérique serait le dernier coup tordu, plus traître que les ombres de la caverne, plus trompeur que ne l'a jamais été un dispositif d'illusion avant lui. Et, alors que le

¹ *Ibid.*, p. 63., Traduction personnelle : « The structure of a computer image is a case in point. On the level of representation, it belongs to the side of human culture, automatically entering in dialog with other images, other cultural "sems" and "mythemes." But on another level, it is a computer file which consist from a machine-readable header, followed by numbers representing RGB values of its pixels. On this level it enters into a dialog with other computer files. The dimensions of this dialog are not the image's content, meanings or formal qualities, but file size, file type, type of compression used, file format and so on. In short, these dimensions are that of computer's own cosmogony rather than of human culture. »

réalisme immersif des simulations informatiques ne cesse de progresser, comment ne pas reconnaître une raison à ceux qui alertent sur le trompe-l'œil de la « réalité virtuelle » ? Vue sous cet angle, elle n'a en effet de réel que le nom. Mais gare aux jugements hâtifs : il convient, avant de rendre la sentence, de remonter aux origines du virtuel. Aristote, l'un des premiers, en donne une définition dans sa *Métaphysique*¹ à travers la notion de « puissance » :

La notion d'acte que nous proposons peut être élucidée par l'induction, à l'aide d'exemples particuliers, sans qu'on doive chercher à tout définir, mais en se contentant d'apercevoir l'analogie : l'acte sera alors comme l'être qui bâtit est à l'être qui a la faculté de bâtir, l'être éveillé à l'être qui dort, l'être qui voit à celui qui a les yeux fermés mais possède la vue, ce qui a été séparé de la matière à la matière, ce qui est élaboré à ce qui n'est pas élaboré. Donnons le nom d'acte au premier membre de ces diverses relations, l'autre membre, c'est la puissance.²

Chez Aristote, le virtuel n'est pas opposé au réel mais à l'actuel ; plus précisément, c'est de l'actuel en puissance, du possible qui attend son actualisation. Ceci doit nous mener à une conclusion : le virtuel n'est pas opposé au réel, au contraire il en est l'une des dimension. Seulement, c'est un réel en puissance, un réel idéal qui attend sa réalisation en actes. On comprend un peu mieux l'usage de ce qualificatif pour désigner la réalité informatique : l'opposition de la puissance et de l'acte s'est transfigurée en opposition de l'immatérialité du numérique à la matérialité du monde physique. Voici donc une première clarification : le virtuel numérique n'est pas opposé au réel, c'est même une de ses catégories. Mais la place qu'il occupe dans le royaume du réel paraît encore bien fragile. « Exister » signifie étymologiquement « sortir de », « se manifester », « se montrer » ; en un mot, exister c'est être en actes. L'existence des objets virtuels numérique est donc toujours menacée, sujette à caution. Pourtant, la réalité numérique semble bien différente de la simple puissance aristotélicienne. Qu'on y pense : les images numériques n'existent pas au seul état mental, dans l'esprit d'un joueur de jeu vidéo, mais elles se déroulent bien sur un écran devant ses yeux, et, lorsqu'il actionne le *joystick* de sa manette, il peut constater une réaction à l'écran. Ce virtuel-là comporte en fait de nombreuses caractéristiques de l'actuel aristotélicien du côté duquel on est pourtant bien réticent à le classer. C'est que les objets virtuels numérique sont nimbés d'un mystère ontologique. Une image numérique, par exemple, affichée sur un écran n'a assurément pas la

¹ ARISTOTE, *Métaphysique*, Ed. minor., Paris, J. Vrin, 1991.

² *Métaphysique*, *op. cit.*, livre 8, 6, 1048 a 35-b 5.

corporalité d'une photographie imprimée ou d'un tableau dont on voit à l'œil nu l'enchevêtrement des coups de pinceau – on sent la matière et, sans même la toucher, on est en mesure d'en éprouver la rugosité. Rien de tel pour l'image numérique condamnée, quelle que soit sa nature et son origine, à la platitude lisse et désincarnée de l'écran. Et le problème s'épaissit lorsqu'on remonte une étape en amont, à la source de l'image. L'écran est neutre et ne nous apprend rien – il anonymise en quelque sorte les images – mais il a un corps, une matérialité. Le fonctionnement d'un écran demeure dans le royaume du physique : sous chaque *pixel* logent trois diodes correspondant aux trois couleurs primaires qui s'allument à tour de rôle selon la couleur du *pixel* en question. Mais d'où provient l'ordre qui décide laquelle des trois diodes s'allume ? Que se passe-t-il en amont de l'écran ? En amont de la matière ? D'où vient l'image numérique ? Des circuits électroniques de silicium ? De la liste de 0 et de 1 stockée dans un fichier binaire ? D'une description vectorielle ? L'image numérique souffre d'un défaut de clarté dans son fonctionnement, d'une d'opacité qui dissimule sa « machinerie »¹ et qui est à la source de toutes les confusions. Si l'on s'en tient à la double nature des nouveaux médias évoquée par Lev Manovich — existence phénoménale au niveau de l'apparaître, existence binaire au niveau de l'« ontologie de l'ordinateur » — l'image numérique est, en dernier instance, une certaine quantité de *bits*. En ne retenant que ce niveau fondamental, on peut alors s'accorder au moins sur l'existence d'« objets numériques » désignant, au choix, soit un fichier informatique codant une séquence binaire, soit sa manifestation électronique à travers l'activation des transistors correspondants. Mais qu'en est-il des « objets virtuels » constitués à la fois d'un « objet numérique » et de sa manière de nous apparaître ? Une voiture virtuelle dans un jeu vidéo existe-t-elle comme « objet virtuel », au-delà de son existence numérique en tant que fichier binaire ? C'est cette question qui est au cœur du problème ; elle concerne d'ailleurs, au-delà des objets informatiques, les artefacts en général dont le statut ontologique a souffert de remises en cause constantes au cours de l'histoire de la philosophie.

L'un des paradoxes classiques dans la question de l'ontologie des artefacts est celui de la statue d'argile : le bloc d'argile précède l'existence de la statue et celle-ci n'est composée de rien d'autre que de l'argile. Dès lors, la statue existe-t-elle ou bien n'y a-t-il encore et toujours qu'un seul et même bloc d'argile ? Contre le primat ontologique accordé aux objets naturels sur les objets artificiels, la philosophe américaine Lynn Ruder Baker propose, dans son livre *The Metaphysics of*

¹ Expression du philosophe de la technologie Albert Borgmann

*Everyday Life*¹, une thèse originale pour la défense des artefacts, la « constitution view » qu'elle résume de cette manière : « Lorsqu'une chose d'un genre premier se trouve dans certaines circonstances, une chose d'un autre genre premier – une nouvelle chose, dotée de nouveaux pouvoirs causaux – vient à exister. »². Plus généralement, la *constitution view* correspond à l'idée que tout objet dans le monde est constitué, d'abord par d'autres objets, et, en dernière instance, par des agrégats de particules subatomiques. Dès que ces agrégats de particules se trouvent être dans des « circonstances favorables », i. e. dès qu'ils sont arrangés d'une certaine manière et occupent un certain espace à un certain moment, alors une nouvelle chose douée d'un statut ontologique entier apparaît dans le monde. Mais il reste une difficulté à éviter : puisqu'un objet constitué occupe toujours l'exact même place que les agrégats qui le constituent à l'exact même moment, ce qui sembler relier l'objet constitué et ses constituant est une relation d'identité. On retombe sur le paradoxe de la statue qui n'est en réalité que le bloc d'argile. Mais Baker est claire sur le sujet : la *constitution view* ne décrit pas une relation d'identité mais plutôt une « forte unité ». La philosophe prend différents exemples comme celui d'un drapeau rapiécé ou d'une rivière qui s'écoule, chacun de ces objets persistant dans son existence malgré le renouvellement constant des agrégats qui le composent. Finalement, selon la thèse de Baker, chaque objet possède essentiellement une « primary kind », qu'on pourrait traduire par « genre premier », qui détermine ses conditions de persistance. Tant qu'un objet possède son « genre premier » il continue d'exister et disparaît dès lors qu'il le perd. Si Baker emprunte cette notion de *primary kind* à Aristote, à la différence de celui-ci elle l'accorde aux artefacts tout autant qu'aux objets naturels. En effet, puisque selon la *constitution view*, chaque objet du monde est en dernière instance composé d'un agrégat de matière, il n'y a pas de raison de refuser aux objets artificiels le statut ontologique accordé aux objets naturels. Baker prend l'exemple d'un bateau constitué d'un agrégat de planches et de clous :

Chacun des éléments composant l'ensemble des planches et des clous est en soi un artefact. Une planche est constituée d'un ensemble de molécules de cellulose et un clou est constitué d'un ensemble d'atomes de fer. Ainsi, l'ensemble des planches et des clous est lui-même constitué d'un ensemble d'éléments naturels non artificiels : des molécules de cellulose et des atomes de fer. Et ainsi de suite jusqu'aux agrégats de particules subatomiques. Bien que les planches et les clous (ainsi que le bateau) soient des artefacts, les planches et les clous sont constitués d'agrégats d'objets naturels. La constitution ne fait donc pas de distinction entre les artefacts et les non-artefacts (objets

¹ BAKER Lynne Rudder, *The metaphysics of everyday life: an essay in practical realism*, Cambridge, Cambridge university press, « Cambridge studies in philosophy », 2007.

² *Ibid.*, p. 32., Traduction personnelle : « when a thing of one primary kind is in certain circumstances, a thing of another primary kind – a new thing, with new causal powers – comes to exist »

naturels). La relation de constitution s'applique entre les artefacts, entre les artefacts et les non-artefacts, et entre les non-artefacts.¹

Baker oppose une défense solide des artefacts à leur discrimination ontologique et les place au même niveau que les objets naturels : leur existence rend le monde « ontologiquement plus riche ». Si elle ne traite pas directement des objets virtuels numérique dans son livre, sa thèse peut leur être étendue avec le « réalisme virtuel » que le philosophe australien David Chalmers décrit en ces termes : « Les objets virtuels existent réellement et sont des objets numériques. »². Les philosophes français Guillaume Bucchioni et Alexandre Declos décomposent le « digitalism » — la nature numérique profonde des objets virtuels — en s'appuyant sur la *constitution view* de Baker :

table virtuelle ← agrégat (plateau virtuel + pieds virtuels) ← structures de données de haut niveau (code source) ← structures de données de niveau inférieur (code assembleur) ← structures de données de niveau le plus bas (code machine/tableaux de bits) ← agrégats de bits³

Cette définition des objets virtuels est connue sous le nom de « strong digitalism » : les objets virtuels sont en tout et pour tout constitués uniquement d'agrégats de *bits*. Mais cette thèse n'est pas vraiment satisfaisante : elle échoue à rendre compte de la dimension intentionnelle des objets virtuels. Sans aller jusqu'à défendre une « mind-dependence » qui n'aurait pas de sens puisqu'un programme continue d'exister lorsqu'il n'est pas utilisé, force est de reconnaître que de tels objets ne sont pas seulement des agrégats de *bits* mais aussi des manières d'apparaître, d'être conçu et utilisé. Après avoir un temps défendu ce *strong digitalism*, David Chalmers l'amende donc pour prendre en compte la dimension intentionnelle des objets virtuels :

Une version claire du digitalisme virtuel pourrait [...] [dire] que les objets virtuels sont entièrement constitués de structures de données ou fondés sur celles-ci. En fait, ce n'est pas mon point de vue [...] Les mondes virtuels ne concernent pas seulement les structures de données, mais aussi les utilisateurs humains, et de nombreux objets et propriétés virtuels sont en partie fondés dans l'esprit

¹ *Ibid.*, p. 51. Traduction personnelle : « Each of the items in the aggregate of planks and nails is itself an artifact. A plank is constituted by an aggregate of cellulose molecules and a nail is constituted by an aggregate of iron atoms. So, the aggregate of the planks and nails is itself constituted by an aggregate of natural nonartifactual things: cellulose molecules and iron atoms. And so on down to aggregates of subatomic particles. Although planks and nails (as well as the boat) are artifacts, the planks and nails are constituted by aggregates of natural objects. So, constitution does not distinguish between artifacts and non-artifacts (natural objects). The constitution relation holds between artifacts, between artifacts and nonartifacts, and between nonartifacts. »

² CHALMERS David, « The Virtual and the Real », in *Disputatio*, vol. 9, mars 2018, p. 309-352., Traduction personnelle : « Virtual objects really exist and are digital objects »

³ BUCCHIONI Guillaume et DECLOS Alexandre, *The Constitution of Virtual Objects*, Collège de France, s.n., 2023.

des utilisateurs humains. [...] Je suggère ce qui suit : les objets et propriétés virtuels sont fondés sur des structures de données et des objets et propriétés mentaux.¹

À partir de ce nouveau digitalisme, Buccioni et Declos réinterprètent la formule centrale de la *constitution view* : « Lorsqu'un objet numérique d'un certain genre premier se trouve dans certaines circonstances favorables, un objet d'un autre genre premier – un objet virtuel doté de nouveaux pouvoirs causaux – vient à exister »². Ainsi, lorsqu'un certain agrégat de *bits* se trouve dans des circonstances computationnelles et intentionnelles favorables un nouvel objet virtuel vient à l'existence. David Chalmers, dans son livre *Reality* +³ prend l'exemple d'un chaton virtuel qui n'est pas un faux chaton, mais un « vrai chaton virtuel » dont le « genre premier » est « chaton virtuel ». Finalement, la constitution d'un objet virtuel fait une différence ontologique de la même manière qu'un artefact physique ou qu'un objet naturel : le monde devient ontologiquement plus riche.

Maintenant que l'on sait les objets virtuels dotés d'un statut ontologique entier, il nous reste encore à lever le voile sur le sens précis de ce terme de « virtuel ». Chalmers distingue d'abord une conception directement héritée de l'appréciation aristotélicienne : le virtuel en tant que « as if reality ». C'est cette conception qui est à l'œuvre dans la description que fait Anthonin Artaud du théâtre, « une réalité fictive : un monde alternatif qui est une illusion ou un mirage, mais qui possède néanmoins de grands pouvoirs »⁴. Cette acception correspond, selon Chalmers, au « fictionalisme virtuel » qui reconnaît une existence réelle aux objets numériques, mais seulement fictive à leurs correspondants virtuels. Si cette conception a une part de validité lorsqu'on traite d'un jeu vidéo ou d'un film d'animation incluant, bien sûr, une dimension fictive, elle ne doit pas pour autant nous tromper. Contre le « fictionnalisme virtuel » qui ne reconnaît une existence qu'à l'objet numérique tout en renvoyant l'objet virtuel au statut de fiction, Chalmers oppose le « réalisme virtuel » qui, comme on l'a vu, reconnaît l'existence pleine et entière de l'objet virtuel en tant que constitué d'un objet numérique et d'états mentaux. Rompant avec la compréhension héritée d'Aristote d'un virtuel en puissance, fictif, « as if », le virtuel numérique est bien plutôt du côté de

¹ CHALMERS David, « The Virtual as the Digital », in *Disputatio*, vol. 11, 1 décembre 2019, p. 453-486., Traduction personnelle : « One clear version of virtual digitalism might ... [say] that virtual objects are wholly constituted by, or grounded in data structures. In fact this is not my view [...] Virtual worlds involve not just data structures but also human users, and many virtual objects and virtual properties are grounded in part in the minds of the human users. [...] I suggest the following: virtual objects and properties are grounded in data structures and mental objects and properties »

² G. Buccioni et A. Declos, « The Constitution of Virtual Objects », art cit.

³ CHALMERS David, *Reality+ : virtual worlds and the problems of philosophy*, First edition., New York, NY, W. W. Norton & Company, 2022.

⁴ *Ibid.*, p. 192. Traduction personnelle : « an as-if reality: an alternative world that's an illusion or mirage but nevertheless has great powers. »

l'actuel, de la réalité qui agit, qui existe et fait une différence concrète dans le monde. C'est cette définition, la seule qui vaille concernant le numérique, que retient Stéphane Vial dans *L'être et l'écran* :

Le virtuel informatique, c'est donc le simulationnel au sens technique du terme, c'est-à-dire en tant que résultat d'une manipulation programmable de l'information. Il ne faut pas le confondre avec le simulacre, qui nous ramènerait dans la caverne platonicienne des mensonges et des mirages, réduisant la technicité objective du virtuel à une métaphysique fantasmée de l'illusion. Le simulationnel, au contraire, est entièrement réel : on en relève des applications concrètes dans l'efficacité opérationnelle du simulateur de vol, la précision scientifique du logiciel de conception assistée par ordinateur, le réalisme incroyable du jeu vidéo. Le plus souvent ignorée, en particulier par tous ceux qui aiment la confondre avec quelque métaphysique de l'irréel, cette acception strictement technique du terme « virtuel » est la seule qui soit objectivement recevable, et par conséquent la seule que nous retiendrons.¹

Le virtuel auquel on fait référence lorsqu'on parle des simulations de l'informatique graphique, c'est donc le « virtuel informatique » ne devant son qualificatif de « virtuel » qu'à sa programmabilité et non à une quelconque irréalité ou inactualité. Ce virtuel n'est pas une pure idée : il lui faut tout un appareillage matériel pour fonctionner, il se manifeste visuellement sur un écran et permet l'action. Fort de cette nouvelle définition, le virtuel peut s'affirmer dans sa dimension simulationnelle, comme un nouveau mode d'imitation de la nature à la suite de la représentation picturale et de la reproduction photographique.

De la reproduction à la simulation

Dans le but de saisir pleinement la programmabilité des nouveaux medias, Manovich appelle à passer des « medias studies » au « software studies » : les objets virtuels prennent bien souvent part à des simulations, processus de création mystérieux empruntant autant à la représentation qu'à la reproduction. À la représentation, car la simulation est programmée par un concepteur humain qui se sert du code comme un peintre se sert de ses pinceaux pour retranscrire une réalité qui ne se départ jamais complètement d'une certaine subjectivité. À la reproduction, car la simulation génère son contenu de manière automatique — au moment même où elle « tourne », elle n'obéit qu'aux lois de son programme, comme l'image photographique n'obéit qu'aux lois

¹ S. Vial, *L'être et l'écran. Comment le numérique change la perception*, op. cit., p. 159.

physiques et chimiques à l'instant où la lumière pénètre la chambre noire pour s'imprimer sur la pellicule. La simulation constitue donc une sorte de représentation automatisée qui, bien que sa réalité ne découle pas d'un contact immédiat avec le monde physique à l'instar de la photographie, peut prétendre à une forme d'objectivité du fait de cette automatisation. On l'a vu plus haut, l'ordinateur est une machine qui, comme toutes les machines, met en œuvre un mécanisme causal. Dans un programme informatique, l'ordre des lignes de codes, du haut vers le bas, importe car chacune d'entre elles constitue une étape logique découlant des précédentes et influençant les suivantes. Ainsi, le programmeur occupe-t-il une position à part vis-à-vis du peintre et du photographe. Il possède la liberté de la peinture mais l'automatisme de la photographie : au lieu d'intervenir directement sur chaque détail du monde qu'il cherche à représenter, il décrit plutôt les lois générales qui vont le faire advenir. La simulation informatique réalise en cela, à la perfection, la vision mécaniste d'un monde qui tourne comme une horloge bien huilée, réglé une fois pour toute à son commencement par un grand horloger. Seulement, l'horloger n'est plus ici Dieu mais le programmeur, et les lois physiques sont remplacées par les lignes de code. L'informatique s'inscrit dans une continuité parfaite de l'idéal de la figuration naturaliste que poursuit la Modernité, et dont elle constitue la dernière actualité. Après la mise au point de la perspective, et celle de l'appareil photographique, la simulation informatique apparaît comme un nouveau dispositif naturaliste né du concours de l'art et de la science pour aller toujours plus loin, ou plutôt, plus en profondeur, dans la capture du réel. En effet, là où la photographie ne reproduisait que la « surface » des phénomènes, leur apparence visuelle, la simulation, fidèle à la double nature des objets virtuels, vise une reproduction, non seulement de la surface, mais aussi de la profondeur du réel : son fonctionnement. L'idéal naturaliste est poussé bien plus loin : il s'agit de reproduire le monde dans toute sa pureté, à partir de ses seuls principes mathématiques et des règles physiques qui le gouvernent. Ainsi, contre l'idée classique qui l'associe à l'illusion, il convient de reconnaître l'importance de la simulation en tant que dispositif permettant d'accéder au réel dans sa dimension systémique, celle-là même qui est au cœur du paradigme mécaniste.

L'usage scientifique des simulations constitue un exemple frappant de cette reconnaissance. Les scientifiques, à la poursuite, depuis l'avènement de la science moderne, des dispositifs expérimentaux les plus objectifs possibles pour donner corps aux vérités scientifiques, s'emparent désormais de la simulation pour fabriquer des faits. Si ce nouvel outil n'a pas l'autorité de véritables observations empiriques, il constitue un troisième terme intermédiaire entre hypothèses et faits observables. La simulation se révèle particulièrement utile pour tester les prédictions empiriques

d'une hypothèse trop complexe, voire impossible à mettre en œuvre dans une expérience physique. C'est le cas, par exemple, des hypothèses concernant les phénomènes évolutifs comme l'apparition et le développement du langage. Dans leur ouvrage *Simulating the evolution of language*¹, Angelo Cangelosi et Domenico Parisi défendent les avantages de la simulation informatique pour faire avancer leur champ de recherche :

Les simulations sont des outils permettant de développer divers « expériences ». Elles peuvent être considérées comme des laboratoires expérimentaux virtuels qui permettent aux chercheurs de mener des expériences réalistes, impossibles et contrefactuelles. Les simulations peuvent être utilisées pour reproduire empiriquement des expériences de laboratoire réalistes, avec l'avantage supplémentaire que la manipulation des conditions et des variables est plus facile et peut être étendue à des conditions et des variables qui ne peuvent pas être manipulées dans des expériences réelles. [...] Les simulations informatiques permettent au chercheur de créer des scénarios dans lesquels un grand nombre d'entités se voient attribuer un ensemble de propriétés et où l'ordinateur fait interagir ces entités de manière à ce que le chercheur puisse observer (littéralement, voir sur l'écran de l'ordinateur) les propriétés globales qui émergent de leurs interactions.²

Le principe de la simulation, et ce qui fait son intérêt dans un usage scientifique, est finalement très proche de l'ambition photographique : mettre la subjectivité humaine de côté. Ici, le scientifique se contente de régler les paramètres de base avant de lancer la simulation ; il peut alors se retirer au rang de simple observateur comme il le ferait vis-à-vis d'une expérience physique.

Concernant les images, la synthèse 3D repose également sur une véritable simulation de mondes virtuels. Ces « nouvelles » images, bien que toujours affichées sur un écran, se distinguent en ce qu'elles sont générées à partir de données géométriques en trois dimensions. En plus des axes x et y des abscisses et des ordonnées, l'axe z de la profondeur est désormais pris en compte. Avec une telle représentation spatiale, de nouvelles possibilités s'offrent à l'informatique graphique : il devient possible d'appliquer les lois de la physique au calcul des images, à commencer par celles qui régissent la lumière. En 1968, Arthur Appel met au point le premier algorithme de *ray casting* qui consiste à simuler la distribution des rayons lumineux à travers une scène 3D. La lumière jaillit dans les mondes virtuels : après la photographie, la synthèse 3D se sert de la physique pour produire

¹ CANGELOSI Angelo, *Simulating the Evolution of Language*, 1st ed., London, Praxis, 2002, 1 p.

² *Ibid.*, p. 8-12., Traduction personnelle : « Simulations are tools for the development of a variety of "experiments". They can be considered as virtual experimental laboratories that allow researchers to run realistic, impossible, and counterfactual experiments. Simulations can be used to replicate empirically realistic laboratory experiments with the added advantage that manipulation of conditions and variables is easier and can be extended to conditions and variables that cannot be manipulated in real experiments. [...] Computer simulations allow the researcher to create scenarios in which a large number of entities are assigned a set of properties and the computer causes these entities to interact in such a way that the researcher can observe (literally, see on the computer screen) the global properties that emerge from their interactions. »

des images. Au cours des années 1970, l'ingénieur français Henry Gouraud ainsi que les ingénieurs américains et vietnamiens, Jim Blinn et Bui Tuong Phong, posent les bases du *shading*, c'est-à-dire la simulation des matériaux selon leurs propriétés physiques. Les découvertes se succèdent dans le domaine, jusqu'à la mise au point de la *rendering equation* en 1986 par David Immel et James Kajiya qui constitue un tournant majeur dans la recherche. Cet algorithme, encore utilisé aujourd'hui, définit précisément la façon dont chaque rayon de lumière virtuel doit se comporter selon les propriétés physiques de la surface qu'il heurte, permettant ainsi d'obtenir une image 2D rastérisée à partir d'une scène 3D virtuelle : chaque *pixel* de l'image rendue est défini en fonction de l'angle de vue de la caméra, de la quantité de lumière et des objets présents dans la scène. Bien que l'équation de rendu soit restée inchangée, les progrès techniques améliorent, année après année, la précision du calcul. Parmi ces avancées, la mise au point des cartes graphiques marque un nouveau pas vers le réalisme simulationnel en permettant l'avènement des jeux vidéo en 3D. Le philosophe Mathieu Triclot parle de l'arrivée de cette troisième dimension dans le jeu vidéo comme d'une « révolution aussi importante que le passage du muet au parlant pour le cinéma »¹ et ajoute : « avec la 3D c'est le jeu lui-même qui se mue dans son ensemble en un genre de quasi-cinéma »². Cette nouvelle technologie combine l'interactivité des jeux vidéo 2D au réalisme physique et spatial des images 3D :

Le rendu en temps réel consiste à créer rapidement des images sur l'ordinateur. C'est le domaine le plus interactif de l'informatique graphique. Une image apparaît à l'écran, le spectateur agit ou réagit, et ce retour d'information affecte ce qui est généré ensuite. Ce cycle de réaction et de rendu se produit à un rythme suffisamment rapide pour que le spectateur ne voie pas des images individuelles, mais soit plutôt immergé dans un processus dynamique.³

Jusqu'à là, les images de synthèse étaient calculées selon un mode de rendu dit « offline », nécessitant plusieurs heures, voire plusieurs jours, de calcul pour une seule image. À ce rythme, l'interactivité était impossible et seul le cinéma pouvait faire un usage de cette technologie. Avec

¹ TRICLOT Mathieu, *Philosophie des jeux vidéo*, Paris, la Découverte, « La Découverte-poche », 2017, p. 87.

² *Ibid.*, p. 88.

³ AKENINE-MÖLLER Tomas, HAINES Eric, HOFFMAN Naty, et al., *Real-time rendering*, Fourth edition., Boca Raton London New York, CRC Press, Taylor & Francis Group, « An A K Peters book », 2018, 1178 p, p. 1., Traduction personnelle de : « Real-time rendering is concerned with rapidly making images on the computer. It is the most highly interactive area of computer graphics. An image appears on the screen, the viewer acts or reacts, and this feedback affects what is generated next. This cycle of reaction and rendering happens at a rapid enough rate that the viewer does not see individual images, but rather becomes immersed in a dynamic process. »

l'arrivée de la *GPU*, la *Graphic Processing Unit*, vient le rendu « realtime », en temps réel : chaque image est désormais calculée dans la fraction de seconde qui précède son affichage à l'écran, lui permettant de réagir aux actions de l'utilisateur. Les consoles de jeu comme la *Sony Playstation*, la *Sega Saturn* ou la *Nintendo 64*, se dotent des capacités de rendu 3D en temps-réel, et puis c'est au tour des ordinateurs. À la fin des années 1990, la carte graphique *GeForce 256* de *Nvidia* fait son apparition pour les ordinateurs alors que les *API* (*Application programming Interface*) *OpenGL* et *DirectX* sont créées pour répondre aux besoins nouveaux de la programmation 3D. L'arrivée du rendu en temps réel donne enfin une existence concrète à la « réalité virtuelle » que David Chalmers décrit comme « un espace immersif, interactif et généré par ordinateur » ¹. L'immersivité c'est le fait d'être plongé dans un environnement perceptible. C'était déjà l'objectif du cinéma argentique poursuivi avec une intensité nouvelle à l'ère numérique. C'est que désormais, au-delà d'une simple immersivité perceptive, la réalité virtuelle permet aussi l'interactivité. Celle-ci repose sur les caractéristiques de programmabilité déjà évoquées du numérique et implique une relation à double sens : « Interactif signifie qu'il existe une interaction bidirectionnelle entre les utilisateurs et l'environnement, ainsi qu'entre les objets présents dans l'environnement. »². Enfin, la troisième caractéristique de la réalité virtuelle est celle qui entérine sa discontinuité avec l'argentique : elle est générée par ordinateur. Mais, ce n'est pas parce qu'elle existe de manière immatérielle, ou du moins, numérique qu'elle ne peut exister comme un espace, un lieu voire un « monde virtuel » :

Nous tournions autour des images, maintenant nous allons tourner dans les images. [...] Les images virtuelles ne sont jamais seulement des images, juste des images, elles possèdent des dessous, des derrières, des en-deçà et des au-delà, elles forment des mondes. [...] Un monde virtuel est une base de données graphiques interactives, explorable et visualisable en temps réel sous forme d'images de synthèse tridimensionnelles de façon à donner le sentiment d'une immersion dans l'image. Dans ses formes les plus complexes, l'environnement virtuel est un véritable « espace de synthèse », dans lequel on peut avoir le sentiment de se déplacer « physiquement ».³

Une telle conception de la réalité virtuelle semble ouvrir la voie à la notion de « prise de vues virtuelles » qui serait l'enregistrement avec une caméra virtuelle d'un point de vue dans un monde

¹ D. Chalmers, *Reality+*, *op. cit.*, p. 193., Traduction personnelle : « A virtual reality environment is an immersive, interactive, and computer-generated space. »

² *Ibid.*, p. 194., Traduction personnelle : « Interactive means that there's two-way interaction between users and the environment, and among objects in the environment. »

³ QUÉAU Philippe, *Le virtuel: vertus et vertiges*, Paris, Champ Vallon ; Institut national de l'audiovisuel, « Collection Milieux », 1993, 215 p.*Ibid.*

virtuel. La « prise de vues » a ceci de propre qu'elle est l'enregistrement automatique et objectif d'un point de vue dans le monde. Si, dans le cas de la prise de vues réelles, ce monde désignait le monde physique, il ne doit pas y avoir de raison de refuser à un monde virtuel, dès lors qu'il met en œuvre un espace dans lequel se déplacer et une certaine persistance, la capture des points de vue qu'il offre. Avec la prise de vues virtuelle, le « ça a été » de l'image argentique serait donc conservé, simplement il ferait référence à un « ça a été » virtuel. Et de fait, les moteurs de rendu 3D sont de véritable simulation de la prise de vues avec des caméras et des lumières virtuelles calquées sur des propriétés physiques pour un photoréalisme qui fait leur succès.

Au début des années 2000, les images 3D remplacent progressivement la 2D dans de nombreux domaines comme les effets spéciaux et le jeu vidéo. Les *blockbusters* comme *Matrix* ou *Le seigneur des anneaux* reposent largement sur cette technologie ; le film de Peter Jackson allant jusqu'à investir le champ du *real-time* en réalisant la première *motion capture* en temps-réel avec le personnage de Gollum – la technologie mise au point pour le film permettait de visualiser, directement sur le plateau, la créature de synthèse se mouvoir selon le jeu de l'acteur revêtu d'une combinaison de capteurs. Les images 3D atteignent un niveau de réalisme dans la reproduction de personnages humains qui les fait basculer dans la *uncanny valley*, à la lisière de l'illusion parfaite. La chercheuse Marie-Paule Cani, dans son article *Mondes virtuels : genèse des nouveaux mondes*¹ offre un bon aperçu de la prouesse technique que représentent ces nouveaux effets spéciaux basés sur des simulations informatiques :

« Simuler » un monde virtuel en mouvement – avec ses rivières, ses nuages, sa végétation, ses bâtiments, ses personnages, ses animaux – revient potentiellement à y intégrer la quantité colossale de connaissances qui s'y rapportent, issues de tous les domaines scientifiques, ainsi que les valeurs des paramètres permettant de passer du phénomène général aux éléments particuliers du monde observé.²

Dans sa tentative de reproduire le monde, la recherche en informatique graphique se retrouve à devoir comprendre, étudier et inclure toutes les autres sciences. Dès lors, il faut bien mesurer la quantité de connaissances que contient chaque doublure numérique, chaque créature de synthèse, chaque explosion générée par ordinateur. Les effets spéciaux numériques n'ont rien à envier ni en complexité ni en spectacularité aux effets pratiques. Mais bien sûr, ces simulations sont

¹ CANI Marie-Paule, « Mondes virtuels : genèse des nouveaux mondes » in *La Création*, s.l., Odile Jacob, « Le Temps des savoirs », 2005, p. 97-110.

² *Ibid.*

imparfaites : tout comme les images numériques faites de *pixels*, quelle que soit le niveau de leur résolution, ce sont toujours des approximations dont la précision se heurte à la puissance de calcul. Dès lors, il faut ruser, « tricher » avec le réel, viser le « réalisme perceptif ». En un mot, il faut avoir recours au « trompe l'œil » :

Concentrant la puissance de calcul sur la partie vue du monde, ces techniques optent à chaque instant pour la représentation minimale capable de donner l'illusion du réel, en adaptant à notre insu le niveau de détail d'un modèle selon son importance à l'image, voire en le faisant passer d'un mode de fonctionnement à un autre. Les créateurs de mondes virtuels sont donc amenés à re-découvrir le vieux concept de trompe-l'œil, mais ce dernier prend maintenant une dimension dynamique, chaque objet s'adaptant et se transformant au fur et à mesure qu'on s'en approche, et disparaissant dès que nous ne pouvons plus l'observer.¹

C'est un exercice non seulement de synthèse mais de réécriture : pour simuler une chevelure qui flotte dans le vent, un nuage ou une prairie, il n'est pas possible de reproduire à l'identique les connaissances physiques, météorologiques ou biologiques relatives à ces différents phénomènes. Il faut partir de l'observation du réel et prendre ses distances ; d'abord en faisant des approximations, et puis en inventant des techniques nouvelles. Les lois scientifiques doivent être adaptées ou réécrites quand il s'agit de les appliquer aux mondes virtuels. De même, les corps simulés de manière si réaliste qu'ils donnent l'impression d'avoir affaire à des êtres de chair et d'os, des êtres « épais », ne sont en réalité que des enveloppes vides : tout objet virtuel dans un moteur 3D est défini par sa seule surface. On trouve là de quoi nourrir les critiques d'un virtuel coupable de produire des illusions. Pourtant, le cinéma n'a pas attendu les images de synthèse pour « tricher » avec le réel. S'il s'est toujours présenté comme l'enregistrement inaltéré d'un « ça a été », les liens qu'il entretient avec l'animation et ses techniques de trucage sont plus nourris qu'il ne veut bien le laisser entendre. Le cinéma-illusion de Méliès est sans nul doute un art de la simulation, pas de surprise de ce côté. Mais en creusant un peu plus, c'est aussi le cinéma-vérité de Bazin qui pourrait révéler sa dimension simulationnelle. L'autorité de cet art, qui s'était toujours présenté comme l'enregistrement d'un « ça a été », se fonde, toute entière, sur l'idéal de la figuration naturaliste. Avec son caractère indicielle, l'image argentique prétend avoir atteint l'objectivité parfaite dans l'imitation de la nature. Mais est-ce bien vrai ? Les questions ouvertes par le numérique nous

¹ *Ibid.*

invitent à un nouvel examen de la figuration naturaliste. Nous pourrions alors découvrir que les images, loin d'être des objets figés, sont d'abord des expériences, des images en train de se faire.

III. L'expérience de la prise de vues

Contre la figuration naturaliste

La figuration naturaliste cherche à détacher la figuration de la subjectivité. Au moment où l'on découvre les lois universelles du monde mécaniste, la figuration est l'un des champs de bataille où s'affronte subjectivité et objectivité. On cherche à résoudre, voire à effacer le « mystère de l'intérieur ». L'expérience subjective est un leurre qu'il faut tenir à bonne distance. C'est le projet de Descola dans son anthropologie de la figuration que de retracer, entre autres, l'histoire du naturalisme à travers la figuration pour en souligner tous les errements :

Porteur d'une contradiction initiale entre l'universalité affirmée des lois de la nature et la singularité têtue des humains, le naturalisme s'est développé en promouvant une connaissance du monde qui se voulait de plus en plus indemne des préjugés et errements d'un sujet connaissant. La conquête de l'objectivité au prix d'une effacement de l'intériorité humaine s'est manifestée de la façon la plus nette dans les images et c'est au premier chef à travers elles que l'on peut en retracer l'histoire.¹

Cette histoire de la poursuite d'une objectivité parfaite dans les images est déjà grosse d'une contradiction présente en germe dès la première tentative de la figuration naturaliste pour toucher à l'absolu : la perspective. Dans son essai classique *La perspective comme forme symbolique*², l'historien de l'art Erwin Panofsky démontre avec une force de conviction qui a fait date que la perspective n'est qu'un appareillage symbolique pour représenter le monde, loin de l'objectivité qu'elle a voulu incarner. Panofsky revient sur l'origine de cette technique, on s'en souvient née des expérimentations artistiques et scientifiques d'hommes de la Renaissance comme Giovanni Battista della Porta, notamment au moyen de la *camera obscura*. L'historien en formule une première définition :

Pour nous, la perspective, au sens prégnant du terme, est donc l'aptitude à représenter plusieurs objets avec la partie de l'espace dans laquelle ils se trouvent, de telle sorte que la notion de support matériel du tableau se trouvent complètement chassée par la notion de plan transparent, qu'à ce que

¹ P. Descola, *Les formes du visible*, op. cit., p. 518.

² PANOFSKY Erwin, BALLANGÉ Guy et DALAI EMILIANI Marisa, *La perspective comme forme symbolique: et autres essais*, Paris, Les Éditions de Minuit, 1975, 273 p.

nous croyons notre regard traverse pour plonger dans un espace extérieur imaginaire qui contiendrait tous ces objets en apparente enfilade et qui ne serait pas limité mais seulement découpé par les bords du tableau.¹

La volonté des peintres qui développent en premier cette technique est donc bien celle de supprimer la représentation : rendre invisible le tableau, l'écran opaque qui arrête la vue et s'impose comme une image dont les deux dimensions proviennent d'un monde abstrait et imaginaire, construction de l'artiste. Au contraire, il faut que la peinture soit désormais comme une fenêtre ouverte sur un monde qui doit le moins possible trahir son origine imaginaire pour se rapprocher au plus près de l'univers physique. En ce sens le tableau n'est plus un écran faisant obstacle, mais un « plan transparent » qui donne accès à un véritable espace dont on est en mesure d'imaginer la continuité hors du cadre. Ce faisant, le peintre qui parvient à restituer de la manière la plus fidèle possible cet espace objectif devient un « parfait œil-enregistreur », il s'efface pour n'être plus qu'un simple intermédiaire mécanique entre le monde et son image. L'espace du tableau n'est pas né de l'esprit de l'artiste qui l'a peint mais au contraire, il le précède et lui est extérieur ; il n'est ni soutenu par l'esprit du peintre ni par la toile dont il dépasse le cadre. Tout annonce déjà la prise de vues photographique : la capture sur un plan en deux dimensions d'un espace qui existe hors du cadre et surtout hors de l'esprit de l'artiste. Mais si elle utilise la *camera obscura*, la Renaissance n'a pas encore réussi à automatiser l'opération de transmission entre le monde et son image. Elle s'appuie toujours sur des peintres de chair et d'os dont la subjectivité est assurément le point faible du système de figuration. Pour remédier à cette fragilité, le dispositif de la perspective est sans cesse amélioré dans le but de parvenir à la « construction d'un espace entièrement rationnel, c'est-à-dire infini, continu et homogène. »². Ce dispositif doit servir de tuteur à la subjectivité de l'artiste qui sans cesse l'entraîne de travers, il est là pour corriger son regard et le peintre sait qu'il doit privilégier cette vue à celle de ses sens. Panofsky décrit plus en détails ce système :

On présuppose facilement dans toute cette « perspective centrale » deux données essentielles : d'abord que notre vision est le fait d'un œil unique et immobile ; ensuite que le plan d'intersection de la pyramide visuelle peut à juste titre passer pour une reproduction adéquate de l'image visuelle.³

¹ *Ibid.*, p. 39.

² *Ibid.*, p. 41.

³ *Ibid.*, p. 42.

On retrouve là les deux conditions de l'objectivité que poursuit la figuration naturaliste et que semble atteindre finalement la photographie deux siècles après la Renaissance : « l'œil unique et immobile » et la « reproduction adéquate » de ce que voit cet œil absolu. Ces deux éléments sont les deux pièces d'une machine enregistreuse vidée de toute intériorité, surtout privée de corps. Car un tel dispositif ne correspond certainement pas à la réalité perçue de manière « incorporée », point de vue d'un corps dans le monde. Pour la Renaissance, qui ne dispose pas d'une pareille machine et doit par conséquent s'appuyer sur des corps, ceux des artistes, l'objectivité de la « perspective centrale » ne peut être qu'une abstraction, une représentation. Les deux présupposés de « l'œil unique et immobile » et de la « représentation adéquate de l'image visuelle » sont trompeurs ; ils « reviennent en vérité à faire hardiment abstraction de la réalité, s'il nous est permis dans ce cas de désigner par « réalité » l'impression visuelle subjective. »¹. Le peintre de la Renaissance, quand bien même il se doterait de tout l'appareillage technique et conceptuel de la perspective, n'atteindra jamais que cela, une objectivité conceptuelle tout entière soutenue par la force de sa subjectivité. C'est que l'espace rationnel et infini de la perspective va à l'encontre de l'espace perçu :

la structure d'un espace infini, continu et homogène, en un mot d'un espace purement mathématique, se situe très précisément à l'opposé de la structure de l'espace psychophysiologique. [...] L'espace homogène n'est donc jamais un espace donné : c'est un espace engendré par une construction. Ainsi donc le concept géométrique d'homogénéité peut très exactement être exprimé par le postulat selon lequel, à partir de chaque point de l'espace, il est possible d'effectuer des constructions semblables en tous lieux et dans toutes les directions. Dans l'espace de la perception immédiate, ce postulat ne peut jamais être satisfait.²

Panofsky souligne l'impossibilité qu'il y a d'échapper à la primauté de l'espace psychophysiologique. En effet, c'est le cadre le plus grand qui englobe toutes les représentations, c'est la base sur laquelle peuvent se développer des abstractions comme l'espace mathématique de la perspective. Ainsi, cette représentation rationalisée est vouée à être, selon la formule de Panofsky, une « forme symbolique », au fond pas plus objective ni naturaliste que la peinture symbolique de l'art Moderne. Il faut alors que le naturalisme se trouve une nouvelle monture bien plus puissante que la peinture et sa perspective : c'est chose faite avec la photographie.

¹ *Ibid.*

² *Ibid.*

L'espoir est grand, les Modernes possèdent enfin un dispositif de figuration accomplissant, dans le champ visuel, le rêve de cette objectivité absolue si longtemps poursuivie : « Après six siècles d'expérimentation avec les images, le naturalisme semblait être parvenu, en matérialisant l'esprit, à transformer l'ineffable en figurable, comme une empreinte et non plus comme un succédané. »¹. La photographie réalise au-delà des espérances les deux conditions de la représentation objective énoncées par Panofsky. La première, l'« œil unique et immobile », c'est l'œil de la caméra qui, s'il peut se déplacer entre les images, reste toujours immobile par rapport à chacune d'entre elles. En effet, après quelques progrès techniques, le temps de pose devient si court que la photographie capture des instantanés, des atomes de moments. C'est cette instantanéité qui rend possible le cinéma, succession d'immobilités dont le mouvement n'a lieu qu'entre les photogrammes. De l'autre côté, la seconde condition à la représentation objective, la « reproduction adéquate de l'image visuelle » trouve une réalité avec le procédé chimique qui permet de retenir les photons sur la pellicule. Enfin, la représentation se passe d'un intermédiaire subjectif qui s'efforce en vain de se faire oublier ; elle atteint l'idéal naturaliste de l'image indicielle, basculant ainsi dans la reproduction. Panofsky avait sans doute raison en insistant sur le primat de la perception immédiate, de l'espace psychophysiologique, dont tout point de vue incorporé ne peut s'affranchir. Mais la Modernité a fourni une solution qui, sans renier la conclusion de Panofsky, la contourne de la manière la plus simple : en se débarrassant du corps... Le propre de la machine, ici l'appareil photographique, c'est qu'elle n'a pas de perception et qu'elle peut dès lors accéder immédiatement à l'espace mathématique. Si la perspective s'efforçait de faire croire à un espace imaginaire, la photographie montre un espace réel. L'espace du tableau en perspective centrale suinte le symbolisme. La photographie nous donne au contraire une empreinte, son objet même.

Mais est-on si sûr d'être parvenu à ce point de vue désincarné, de « nulle part » ? Si le moment même de la production de l'image photographique, lorsque la lumière pénètre le diaphragme et s'imprime sur la pellicule, est opéré de manière mécanique, l'œil de la machine a en revanche été aligné sur autre œil, humide et chaud, l'œil d'un corps, celui du photographe placé dans le monde. Roland Barthes, dans *La chambre claire*², distingue entre la photographie pour le spectateur, qui correspond à l'idéal objectif de « la révélation chimique de l'objet »³ et la « photographie de l'Operator » ou « Photographie-selon-le-photographe », qui est avant tout un

¹ P. Descola, *Les formes du visible*, op. cit., p. 530.

² R. Barthes, *La chambre claire*, op. cit.

³ *Ibid.*, p. 24.

point de vue incarné et subjectif : « Je pouvais supposer que l'émotion de l'Operator (et partant l'essence de la Photographie-selon-le-Photographe) avait quelque rapport avec le « petit trou » (sténopé) par lequel il regarde, limite, encadre et met en perspective ce qu'il veut « saisir » (surprendre) »¹. Cette photographie du photographe s'intéresse à l'acte même de photographier et non à l'objet fini de la photographie développée. Prise sous cet angle, la photographie est certainement un point de vue, pas celui de l'œil unique et immobile de l'appareil mais celui mouvant, psychophysiologique du photographe qui « doit lutter énormément pour que la Photographie ne soit pas la mort »². L'objectivité de l'objectif n'apparaît plus ici comme le gage d'un accès direct au monde rationnel de l'espace mathématique infini mais plutôt comme l'assurance de préserver au mieux le point de vue du cadreur ; c'est la vérité en tant que fidélité de l'outil qui prolonge la subjectivité de son utilisateur.

Tout comme la perspective de la Renaissance, la photographie du XIXe est tombée dans l'illusion de la figuration naturaliste. Elle avait pourtant de bonne raison de croire s'être débarrassée pour de bon du point de vue incorporé au profit du point de vue de nulle part. Mais, chassé par la porte, le corps est revenu par la fenêtre. La photographie, pas plus que la perspective, ne saurait s'affranchir du point de vue qui est la condition de toute expérience. Croire le contraire, ce serait oublier l'appartenance au monde de tous ceux qui conçoivent des images, les imaginent, les interprètent ou les utilisent :

Emporté par la perspective exaltante de dévoiler dans des images enfin objectives les ressorts les plus secrets de l'esprit humain, le militant du naturalisme accompli n'avait pas assez pris garde que l'intériorité est plus qu'un mécanisme computationnel analogue à une machine de Turing, que le matérialisme exige toujours un point de vue sur ce qu'il rend manifeste, et que celui-ci ne peut être la simple vigie de l'homme intérieur — « Je ne suis pas seulement logé dans mon corps, ainsi qu'un pilote en son navire », selon la célèbre formule de Descartes [...] Écartelé, sur le front de la figuration comme sur ceux des autres modes d'existence, entre l'irréductibilité têtue d'un moi bavard et l'aspiration à tout soumettre aux lois de la nature, le naturalisme ne cesse de creuser sa contradiction.³

¹ *Ibid.*, p. 23.

² *Ibid.*, p. 31.

³ P. Descola, *Les formes du visible*, *op. cit.*, p. 530.

C'est l'objectif de l'anthropologie de Descola, à la suite de celle de Levi-Strauss, que de montrer le caractère situé de la connaissance, comme de la production d'images. Plus que la prise-de-vues-pour-le-spectateur, objet figé qui fait croire à son objectivité absolue, ce qui nous intéresse désormais c'est la *Photographie-selon-le-photographe* indétachable d'une existence dans le monde. Le cartésianisme a placé la conscience hors du monde, il nous faut une philosophie qui la remette au milieu.

Faire monde

Le solipsisme métaphysique auquel mène le dualisme cartésien place la conscience dans un néant hors du monde à partir duquel elle est forcée de concevoir ses objets de manière abstraite et détachée, comme des représentations plutôt que des présentations. Ce point de vue de nulle part, tout à fait semblable à celui que décrit Panofsky à propos de la perspective, amène la conscience à un paradoxe lorsqu'elle se pose la question de la perception : le perçu est à la fois le vécu (en moi) et son objet (extérieur). Comment résoudre la contradiction apparente ? Dans son livre *La perception. Essai sur le sensible*¹, le phénoménologue Renaud Barbaras revient sur diverses tentatives des philosophies modernes pour concilier ce point de vue de nulle part qui pose des objets absolus avec le phénomène de la perception. L'une des propositions est celle de Locke dont l'empirisme s'oppose à la thèse cartésienne des idées innées inscrites en nous par Dieu. Au contraire, pour l'empirisme, la perception est la seule source de la connaissance. Nous percevons d'abord les objets comme une « collection de sensations simples » avant d'en reconstituer la substance dans la sensation uniforme. Ainsi, pour Locke, la substance est-elle toujours une illusion ou, du moins, une sensation subjective s'apparentant à une idée. L'objet absolu, quant à lui, n'est rien d'autre qu'une collection de perceptions discrètes, à l'image des photogrammes du film dont on reconstitue le mouvement mentalement. L'empirisme de Locke postule donc une réalité extérieure persistante, cause de nos sensations. D'un autre côté, l'idéalisme de Berkeley, souvent opposé à l'empirisme, peut en réalité être considéré comme son exagération. Il part du même point — la perception est la source de toute connaissance, selon la célèbre formule de Berkeley « être, c'est être perçu » — mais il va plus loin en supprimant la réalité extérieure. Là où l'empirisme appuyait la perception sur un donné objectif perçu, l'idéalisme règle le paradoxe de la perception, en moi et dans le monde, en supprimant justement le monde extérieur : il n'y a plus d'autre existant

¹ BARBARAS Renaud, *La perception. Essai sur le sensible*, Hatier., Paris, s.n., 1994.

que les perceptions. Enfin, troisième position, l'intellectualisme cartésien s'inscrit contre la perception en posant une connaissance transcendante accessible de manière purement rationnelle par l'esprit. Mais malgré leurs divergences, l'empirisme, l'idéalisme et l'intellectualisme se rejoignent toujours au même endroit : le point de vue désincarné de la conscience hors du monde qui donne toujours des objets en soi, déjà constitués :

Le moment où quelque chose paraît avant qu'il ne soit cristallisé en chose, où la conscience découvre une réalité avant de la posséder, est négligé au profit de la position d'un pur objet. [...] À l'évidence, ces perspectives projettent sur l'expérience perceptive un cadre philosophique déjà constitué au lieu de rechercher au cœur de la perception ce que signifient réalité et objet.¹

La conception de la perception présente dans l'empirisme comme dans l'intellectualisme va de pair avec une conception philosophique de la réalité « comme pure détermination, c'est-à-dire comme objet »². La réalité est déjà-là, déjà constituée, déterminée et non pas en train de se faire. C'est que toutes ces ontologies pensent l'Être du point de vue de la raison suffisante (« pourquoi y a-t-il quelque chose plutôt que rien ? »). Dès lors, l'Être est défini comme ce qui s'oppose au néant dans lequel il risque sans cesse de basculer. Pour résister à l'appel du vide, il doit être pure positivité, pure détermination. Mais cette vision de l'Être témoigne paradoxalement de la position du philosophe dans une sorte de néant à partir duquel il conçoit l'Être : « L'absence d'être par laquelle commence le philosophe est synonyme de son propre exil du monde »³. Il faut libérer la perception de ce carcan philosophique en commençant par remettre le sujet percevant dans le monde.

C'est ce que cherche à faire la phénoménologie qui se définit étymologiquement comme l'étude des phénomènes. Contre le solipsisme métaphysique du cartésianisme, la phénoménologie replace au contraire la conscience au centre du monde en insistant sur le primat du point de vue incarné, condition de toute expérience. Dès lors, cette philosophie a d'abord une portée épistémologique : elle conçoit l'expérience du « je » dans le monde comme la source première de toute connaissance :

¹ *Ibid.*, p. 26.

² *Ibid.*

³ *Ibid.*, p. 28.

Elle fait ressortir le rôle de l'ego, montrant que la connaissance humaine n'est pas l'œuvre d'un intellect agent séparé des êtres humains, mais l'accomplissement et la possession de quelqu'un qui peut dire « je » et qui peut assumer la responsabilité de ce qu'il dit.¹

Ainsi, Husserl, père fondateur de la phénoménologie, poursuit-il la remise en cause de l'ontologie de la raison suffisante qui pose l'être comme fini et immuable, sorti du néant. Du naturalisme, qui conçoit la nature physique « comme la norme de ce qu'il faut entendre par Être. »², découle une téléologie implicite de la perception : nous percevons une chose en soi, déjà-là, et tout ce que nous percevons est dû entièrement à cette chose-là. Face à cette conception, la difficulté de la philosophie de la perception réside dans la reconnaissance d'une autonomie et d'une pré-existence du réel qui ne vide pas complètement la perception de son pouvoir. C'est ce à quoi veut parvenir Husserl par sa défense d'un retour à l'apparaître. Renversant la conclusion du naturalisme, l'être absolu d'Husserl n'est pas l'être naturel mais la conscience. Dès lors, « être ne peut signifier qu'être pour une conscience »³, et l'apparaître devient du même coup une dimension de l'être :

Husserl veut dire qu'un étant quelconque ne peut être pensé comme tel que par référence à ses modes de données subjectifs, c'est-à-dire du point de vue de sa corrélation avec un sujet, ce qui signifie que l'apparaître est une dimension de l'Être.⁴

La méthode phénoménologique vise donc à accéder à cet absolu qu'est la conscience en suspendant « la vie pré-philosophique », l'« attitude naturelle », cette « croyance en l'existence en soi du monde »⁵. Husserl utilise le terme d'« époché » pour désigner la mise-à-distance de l'attitude naturelle permettant une interrogation du sens d'être de la réalité. Il ne s'agit plus de se demander si le monde existe, mais qu'est-ce que signifie exister pour le monde ? Puisque l'être du monde est désormais l'être pour une conscience, il y a un rapport immédiat et constant entre le monde et la conscience, non par des représentations (comme le décrit Descartes) mais par des présentations :

¹ SOKOLOWSKI Robert, *Introduction to phenomenology*, Cambridge (GB) New York Melbourne, Cambridge university press, 2000, p. 154.

² R. Barbaras, *La perception. Essai sur le sensible*, op. cit., p. 29.

³ *Ibid.*, p. 32.

⁴ *Ibid.*, p. 29.

⁵ *Ibid.*, p. 32.

En reconduisant l'être du monde à la conscience, l'époché invalide définitivement l'ontologie de la vie naturelle et, a fortiori, du naturalisme. [...] Le but de Husserl [...] est de montrer que l'être même du monde consiste à être pour une conscience, sans se dégrader en représentation dans la conscience, c'est-à-dire sans cesser d'être un monde dans la plénitude de son sens.¹

Chez Husserl, et dans la phénoménologie en général, il y a donc une primauté du sensible. Si l'intelligible a toujours droit de cité, et même possède une certaine positivité, il n'est pour autant « pas autonome ; il s'engendre sur la base du sensible qui le fonde ».²

Cet être pour la conscience, cadre globale de toute expérience, passe chez Husserl par l'intentionnalité. Cette notion fondamentale, à l'origine de la phénoménologie, est conçue comme une réponse au solipsisme cartésien. C'est par l'intentionnalité que la conscience retrouve le monde, va à sa rencontre. En effet, pour Husserl, l'idée d'une conscience flottante au milieu d'un néant n'a pas de sens puisque toute conscience est nécessairement « conscience de... », de même que toute expérience est « expérience de... ». Avec l'intentionnalité, la phénoménologie s'inscrit contre la connaissance non située : il ne saurait y avoir d'expérience en générale nous révélant des objets universels, l'expérience de tout phénomène est toujours l'expérience de *ce phénomène-là*. Husserl parle de « noème » pour désigner l'objet intuitionné, et de « noèse » à propos de l'acte même de la perception du noème. Celui-ci n'est pas un objet, il ne doit pas être réifiée comme intermédiaire entre la perception et la chose en soi, au contraire c'est un « moment de la manifestation des choses »³. Ainsi, l'intentionnalité vise le noème sans jamais l'atteindre : il nous est donné par « esquisses ». Du fait de notre expérience située, nous ne percevons jamais un objet de manière absolue, de toutes parts, mais toujours depuis un point de vue particulier. Dès lors, l'objet nous est donné par ses aspects qui se multiplient à mesure qu'on perçoit, qu'on se déplace autour de l'objet et que le temps s'écoule. Ce point de vue ancré de manière spatio-temporelle poursuit l'objet sans jamais l'atteindre de manière absolue. Quel que soit le nombre d'apparitions d'une chose, il y en aura toujours plus en réserve. Mais ces esquisses sont déjà l'objet : dans la phénoménologie, l'apparaître des choses est inclus à leur ontologie. Les apparences sont bien réelles, elles passent du statut psychologique auquel les avait confinées le cartésianisme au statut ontologique. On peut alors comprendre l'*époché* husserlienne comme la mise en parenthèse de nos croyances pour observer les

¹ *Ibid.*, p. 34.

² *Ibid.*, p. 39.

³ R. Sokolowski, *Introduction to phenomenology*, *op. cit.*, p. 48.

intentionnalités à l'œuvre. C'est une manière d'étudier notre être-au-monde en train de se faire, ou comme dirait Descola, notre mondiation. Avec la réduction phénoménologique ontologique qui vise les connaissances scientifiques, Husserl insiste sur le rôle épistémologique de la phénoménologie : montrer que toute connaissance est toujours connaissance d'une conscience. Une fois qu'une science a atteint ce qu'elle pense être une donnée objective, il reste encore à étudier la manière dont se manifeste subjectivement la possession de cette objectivité. On l'a vu, toute science qui se veut purement objective est un « point de vue de nulle part », connaissance de personne. Il faut compléter cette connaissance en la rendant à la conscience intentionnelle. C'est le rôle de la méthode phénoménologique que de pointer l'abstraction des sciences de la nature qui ignorent les intentionnalités sous-jacentes. En ce sens, la réduction phénoménologique est plutôt une augmentation puisque, sans renier l'objectivité scientifique, elle l'enrichit d'un niveau supplémentaire de connaissance en la replaçant dans le jeu des intentionnalités.

L'intentionnalité husserlienne est donc parvenue à ouvrir une voie alternative face à l'impasse dans laquelle s'étaient engouffrés intellectualisme et empirisme. Mais la solution d'Husserl est encore bien fragile. Comme le souligne Barbaras dans son essai, derrière l'intentionnalité qui vise un objet en train d'apparaître, il y a toujours au fond l'idée de l'objet absolu. Même si Husserl insiste sur le fait que cette intentionnalité poursuit un objectif qui recule et lui échappera toujours, force est de constater que cette visée idéale ressemble fortement à la chose en soi. Et de la même manière, à l'autre bout de l'intentionnalité on risque bien de retrouver le sujet cartésien. Le risque de retomber dans l'opposition du sujet et de l'objet est donc toujours bien présent. Dans le dernier chapitre de son ouvrage, Barbaras se propose alors de terminer le travail entamé par Husserl : liquider complètement le sujet cartésien et la chose en soi. Pour ce faire, le phénoménologue propose une ontologie du Sensible contre le positivisme du sujet théorique « capable de prendre une distance infinie vis-à-vis du monde ». Si ce positivisme laisse une place à l'indétermination, elle est vue négativement comme une limite à la connaissance (l'Être étant déterminé). Selon cette idée, ma finitude, en me faisant percevoir le sensible comme un apparaître indéterminé, ne me donne qu'un accès très partiel à la connaissance. Contre cette conception, l'idée d'un sentir comme unité originaire, contact immédiat avec le monde remet fortement en cause l'ontologie classique de la raison suffisante. Puisque nous rencontrons la réalité même dans son apparaître, nous n'avons pas besoin de nous constituer en sujet théorique, hors du monde, pour y avoir accès. L'Être comme Sensible est une absence au cœur de la présence, une fuite, un mouvement permanent — à l'inverse de l'être comme identité. Dans l'ontologie du Sensible, nous

appartenons alors au monde. Notre finitude n'est plus considérée comme une limite mais comme une condition inévitable de l'Être :

Cette philosophie débouche en effet sur le refus de tout surplomb vis-à-vis du monde ; elle est conscience de notre appartenance. Dire que l'Être est sensible, c'est intégrer notre finitude à la définition de l'Être et, par là même, renoncer à la saisir comme limitation.¹

Le Sensible échappe complètement au principe de raison suffisante, puisqu'il ne provient pas du néant, n'en est pas entouré. En revanche, il peut l'intégrer en lui, dans le sens où son Être est indéterminé, refus de positivité. Barbaras distingue alors deux philosophies. La première, qu'il nomme « philosophie de la vision » désigne celle du sujet théorique qui « prend vis-à-vis du monde une distance qui lui permet de le posséder »². Elle conçoit un réel positif, pure présence entourée de néant. La seconde, « philosophie du toucher » s'appuie sur la métaphore tactile pour désigner le contact, la proximité tangible et pourtant l'impossibilité de coïncider avec ce qu'on touche. Ici le sujet percevant appartient au monde, et toute activité qui cherche à dépasser la perception sensible (comme l'exercice de la raison) est certes possible mais toujours conditionnée par cette appartenance première:

Une telle philosophie ne veut certes pas tout réduire à la perception mais elle prétend néanmoins que toute activité qui la dépasse, en particulier l'activité rationnelle, porte la marque de cette appartenance originaire [au monde], et, surtout, que la philosophie doit être la prise de conscience la plus radicale de cette condition.³

Finalement, pour évacuer le sujet du cogito et la chose en soi, Renaud Barbaras rédéfini les termes de la phénoménologie en pensant le sujet de la perception « comme vivant » et le perçu « comme monde »⁴. Cette description rend mieux justice à la manière, incarnée et engageante, dont nous *faisons monde*.

¹ R. Barbaras, *La perception. Essai sur le sensible*, op. cit., p. 75.

² *Ibid.*, p. 77.

³ *Ibid.*

⁴ *Ibid.*, p. 54.

Faire image

Renaud Barbaras pense son ontologie du Sensible comme une manière de renouer avec le monde, de faire corps avec lui. L'analogie du toucher qu'il utilise est révélatrice de notre être-au-monde incarné, incorporé. Nos sens, en contact permanent avec le monde ne coïncident pourtant jamais avec lui, tout comme l'objet que l'on touche. Ce rapport au monde n'est pas sans rappeler la description qu'en fait John Dewey à travers son concept d'« expérience ». Dans son livre *Art As Experience*¹, le philosophe américain décrit l'expérience de la vie d'abord comme l'expérience d'un vivant dans son milieu. Ce dialogue entre la créature et son environnement joue un rôle fondamental dans le développement de l'un comme de l'autre. Ainsi, la vie s'apparente-t-elle ici à une sorte de dialectique nourrie par la suite des équilibres, déséquilibres et rééquilibres de l'être vivant par rapport au monde : « La vie elle-même se compose de phases au cours desquelles l'organisme se désynchronise par rapport au rythme des choses qui l'entourent, puis retrouve son harmonie avec celui-ci. »². Cette succession d'états métastables constitue le rythme de la vie et la base à partir de laquelle peuvent jaillir des harmonies. Dans le flot continue et bruyant de l'expérience en général, prise comme un grand continuum allant de la naissance à la mort, se détachent en effet des expériences singulières qui sont comme des cycles allant d'une tension à une résolution, des moments qui constituent de petites unités. Lorsque s'achève un mouvement parce qu'il a atteint sa fin, parce qu'il est une « consommation et non une cessation »³, on peut le ressaisir dans sa globalité comme une expérience d'un autre genre, une « expérience esthétique ». La sensation d'achèvement et d'harmonie qu'elle procure constitue le matériau de base de l'artiste. Si le terme d'esthétique est souvent confondu avec la perception visuelle, il renvoie plutôt à la perception sensorielle en générale. De ce fait, loin d'être purement conceptuelle, l'expérience esthétique est avant tout l'expérience de la vie. Contre l'idée d'un art détaché du monde, Dewey défend le potentiel esthétique de la vie quotidienne ponctuées de ces petites harmonies :

Les sources de l'art dans l'expérience humaine seront connues de celui qui perçoit comment la grâce alerte du joueur de ballon gagne la foule des spectateurs, qui remarque le plaisir que ressent la ménagère en s'occupant de ses plantes, la concentration dont fait preuve son mari en entretenant le

¹ DEWEY John, *Art as experience*, Perigee trade paperback edition., New York, Berkley Publishing Group, 2005.

² *Ibid.*, p. 20., Traduction personnelle : « Life itself consists of phases in which the organism falls out of step with the march of surrounding things and then recovers unison with it. »

³ *Ibid.*, p. 40.

carré de gazon devant la maison, l'enthousiasme avec lequel l'homme assis près du feu tisonne le bois qui brûle dans l'âtre et regarde les flammes qui s'élancent et les morceaux de charbon qui se désagrègent.¹

Chez Dewey, l'art est donc conçu comme une mise en valeur de l'expérience esthétique, une exacerbation du sentiment de vie qui est comme un pic montagneux se détachant sur l'horizon de l'expérience commune. Les expériences esthétiques n'ont de chance de se produire que dans un monde rythmé, en constante dynamique, en mouvement : vivant. Cette vie, c'est d'abord celle de l'artiste qui perçoit et conçoit depuis son corps plongé dans son milieu. Ainsi, le degré d'éloignement d'une œuvre par rapport à la nature ne tient toujours qu'en apparence : la peinture de Miró nous semble a priori plus éloignée de la nature que celle de Michel-Ange et pourtant Miró a vécu comme Michel-Ange ; comme lui il a respiré et senti, et c'est de son expérience de vie que Miró tire son œuvre, pas moins abstraite en fin de compte que celle du peintre italien. L'artiste est une sorte de catalyseur : luttant contre la dérive du flot continu, il distingue, isole et magnifie les formes de l'expérience esthétique. Leur unité s'inscrit contre l'idée positiviste d'une sensation qui pourrait être décomposée en une juxtaposition de perceptions discrètes.

C'est justement ce présupposé que s'applique à combattre la *Gestalt*, psychologie de la forme. Contre l'empirisme qui voudrait découper l'ensemble d'une mélodie comme la succession des notes qui la composent, elle la conçoit comme une forme que nous percevons immédiatement :

La psychologie de la forme apparaît donc bien comme une récusation du concept empiriste de sensation : nous ne percevons que des formes et, par conséquent le perçu n'est pas composé d'atomes sensibles. [...] toute perception est régie par la structure figure-fond.²

À l'inverse d'un perçu qui serait un « composé d'atomes sensibles », la perception d'une forme est d'abord la perception d'un contraste entre cette forme et le fond. Dès lors, le donné de la perception c'est ce « contraste » qui nous fait voir des images et permet des expériences esthétiques. La *Gestalt* a une grande influence sur la phénoménologie de Merleau-Ponty et on en retrouve des traces dans l'emploi qu'il fait du concept double de Visible et d'Invisible. L'invisible chez Merleau Ponty c'est la structure du monde qui soutient le visible : c'est une « armature intérieure » que le visible

¹ DEWEY John, *L'art comme expérience*, Paris, Gallimard, « Folio », 2010, p. 32.

² R. Barbaras, *La perception. Essai sur le sensible*, op. cit., p. 14.

« manifeste et cache à la fois »¹. Rejouant en cela l'opposition antique du sensible (visible) contre l'intelligible (invisible), Merleau-Ponty l'adapte à la pensée phénoménologique en faisant basculer l'invisible dans le domaine du sensible. Dès lors, « il n'y a pas de monde intelligible, il y a monde sensible. »². L'invisible n'est plus l'opposé du visible, mais il est déjà présent en lui : il en constitue le sens. Dans son étude de Cézanne, Merleau-Ponty voit dans le motif que recherche le peintre cette armature invisible du visible :

Pour tous les gestes qui peu à peu font un tableau, il n'y a qu'un seul motif, c'est le paysage dans sa totalité et dans sa plénitude absolue, — que justement Cézanne appelait un « motif ». Il commençait par découvrir les assises géologiques. Puis il ne bougeait plus et regardait, l'œil dilaté, disait Mme Cézanne. Il « germinait » avec le paysage. Il s'agissait, toute science oubliée, de ressaisir, au moyen de ces sciences, la constitution du paysage comme organisme naissant. Il fallait souder les unes aux autres toutes les vues partielles que le regard prenait, réunir ce qui se disperse par la versatilité des yeux, « joindre les mains errantes de la nature » dit Gasquet. « Il y a une minute du monde qui passe, il faut la peindre dans sa réalité ».³

Le motif que poursuit Cézanne, c'est l'image esthétique déjà présente dans le paysage et qui se révèle au peintre. Chez Merleau Ponty l'artiste ne fabrique pas une image, plutôt il répond à une apparition. À l'inverse de la science qui refuse d'habiter le monde, l'artiste descend sur terre et « prête son corps » : « C'est en prêtant son corps au monde que le peintre change le monde en peinture. »⁴. La perception, en tant que perception d'une forme sur un fond, va dans le sens d'un réel présupposant toujours une subjectivité percevante. « Toute expérience est une reconnaissance »⁵ écrit Barbaras : la perception c'est le regroupement du sensible sous la signification d'une forme. Cette reconnaissance est première par rapport à l'acte rationnel qui divise ensuite l'ensemble en la somme de ses parties. L'argument cartésien du morceau de cire ne tient pas : la cire perçue est d'abord « une harmonie », c'est la connaissance rationnelle qui en extrait des qualités sensibles.

¹ MERLEAU-PONTY Maurice, *Le visible et l'invisible: suivi de Notes de travail*, Paris, Gallimard, « Tel », 2016, 359 p, p. 195.

² *Ibid.*, p. 267.

³ MERLEAU-PONTY Maurice, *Sens et non-sens*, Nachdr., Paris, Gallimard, « Bibliothèque de philosophie », 1996, 225 p, p. 22-23.

⁴ MERLEAU-PONTY Maurice, *L'œil et l'esprit*, Paris, Gallimard, « Collection folio Essais », 1989, 92 p.

⁵ R. Barbaras, *La perception. Essai sur le sensible*, op. cit., p. 17.

En montrant que le réel objectif, visé par l'empirisme comme l'intellectualisme, est une abstraction conçue depuis la réalité vécue, la phénoménologie rejoint la critique anthropologique de Descola. La figuration, comme tous les autres modes de relation au réel, est conditionnée à une situation incarnée : « en effet, on ne voit pas comment un Esprit pourrait peindre »¹. Et, comme dans tous les autres domaines, la pensée positiviste impose un réel unique comme fond de toutes les figurations, instituant de fait une hiérarchie entre les images d'après leur distance plus ou moins grande vis-à-vis de cet étalon de vérité. Dans cette échelle, l'image argentique est tout en haut : seule parmi les images, elle touche le réel. Mais plus pour longtemps : après avoir libéré la peinture de la figuration naturaliste, c'est au tour de la prise de vues d'être libérée par le numérique. À l'heure où toutes les manipulations de *pixels* sont possibles — après *Photoshop*, l'intelligence artificielle porte le coup de grâce — elle n'est plus un critère de vérité. Nous pouvons alors clarifier la situation. D'une part, la remise en cause d'un fond commun qui serait le réel positiviste entraîne dans sa chute son échelle, factice, de la vérité. Désormais, il n'est plus possible de disqualifier une image sur la base de son intensité ontologique ; toutes les figurations, depuis la photographie argentique jusqu'à la peinture en passant par les images de synthèse numériques, sont aussi « réelles » les unes que les autres. D'autre part, puisqu'il n'y a plus de Nature unique, fond sur lequel se déployaient des cultures différentes, mais bien des natures-cultures, coexistent des manières de faire monde non comparables et non mesurables entre elles. De ces mondifications découlent à leur tour des manières tout aussi variées et tous aussi incommensurables de « faire image ». Nous pouvons finalement abandonner une bonne fois pour toutes les querelles ontologiques portant sur des images-objets, pures abstractions, pour nous intéresser à ce « faire image », l'expérience même de la figuration. C'est sous prisme que nous réexaminons la prise de vues.

Un point de vue...

Le « faire image » de la prise de vues se distingue des autres modes de figuration en ce qu'il repose sur la perception. Contrairement à la pensée, au sentiment ou à l'imagination, celle-ci implique la présence concrète de l'objet perçu. La perception est toujours perception de quelque chose. À l'inverse du peintre qui tire les formes qu'il représente de la richesse de son imaginaire, l'opérateur de la caméra est toujours face au monde qu'il enregistre. Si la représentation symbolique vise une absence, la prise de vues est toujours dirigée vers une présence, un apparaître sensible. Ici,

¹ M. Merleau-Ponty, *L'œil et l'esprit*, op. cit., p. 9.

la position du dessin ou de la peinture d'observation est ambiguë. On a vu la description que Merleau-Ponty fait de Cézanne : comme le photographe, l'impressionniste se place en face du monde pour se laisser pénétrer de sa présence. Mais celle-ci est plus une imprégnation, « une germination avec le paysage », qu'une impression dans le sens fort d'une empreinte. Le peintre d'observation, s'il s'appuie en premier lieu sur la perception, laisse toujours entrer une part plus ou moins importante d'abstraction dans sa figuration. À l'inverse, la prise de vues est purement concrète : elle se confond avec la perception sensible du corps, ou plutôt, elle la prolonge. En cela, le *faire image* de la prise de vues est très proche de l'expérience corporelle que décrit Merleau-Ponty et qu'il pose comme condition fondamentale de notre être-au-monde. En effet, « le corps est notre moyen général d'avoir un monde »¹. Mais de cela, on ne doit pas conclure que le corps humain n'est qu'un simple transmetteur qui se contenterait de nous passer l'information absolue d'un réel extérieur, quelle que soit d'ailleurs la provenance de cet extérieur. Ce serait retomber dans le piège de l'empirisme comme de l'idéalisme qui recherchent la nature originelle derrière l'illusion de la sensation, une collection de perceptions sous le monde mouvant des phénomènes, et n'aboutissent en fin de compte qu'à une extériorité morte. C'est cette conception du corps comme simple « transmetteur » que critique Merleau-Ponty dans l'introduction de la *Phénoménologie de la perception*² lorsqu'il décrit « l'appareil sensoriel, tel que la physiologie moderne se le représente » :

Il est inévitable que dans son effort général d'objectivation la science en vienne à se représenter l'organisme humain comme un système physique en présence de stimuli définis eux-mêmes par leurs propriétés physico-chimiques, cherche à reconstruire sur cette base la perception effective et à fermer le cycle de la connaissance scientifique en découvrant les lois selon lesquelles se produit la connaissance elle-même, en fondant une science objective de la subjectivité.³

Le philosophe est très critique de ce corps-interface, fonctionnel, auquel s'oppose justement le projet phénoménologique. Pour lui, « il est inévitable aussi que cette tentative [de reconstruire la perception effective] échoue » car elle cherche à « niveler l'expérience à la hauteur de la nature physique et convertir le corps vivant en une chose sans intérieur. »⁴. Effectivement, c'est bien dans cette optique de remplacer le corps vivant doué d'intériorité par un corps mécanique, pure

¹ MERLEAU-PONTY Maurice, *Phénoménologie de la perception*, s.l., Editions Gallimard, 2013, 637 p, p. 182.

² M. Merleau-Ponty, *Phénoménologie de la perception*, *op. cit.*

³ *Ibid.*, p. 33.

⁴ *Ibid.*

extériorité, que la figuration naturaliste s'est emparée de la photographie. Mais on a vu qu'il est illusoire de chercher à supprimer le point de vue subjectif : « l'œil fixe et immobile » de la caméra est toujours subordonné à l'œil mobile et vivant de son opérateur. Le rapport de force se renverse alors : ce n'est pas le corps vivant qu'on a soumis à l'appareil, c'est au contraire l'appareil qui fonctionne au service du corps comme une prolongation de son point de vue. Notre corps est le « pivot du monde », c'est à partir de lui que nous percevons et que nous agissons : c'est un camp de base à partir duquel nous nous projetons vers le monde. D'une manière générale, la perception élargit le corps : si le corps-objet est limité physiquement, le corps phénoménologique n'a pour limite que ses capacités perceptives, il « excède les limites de la peau »¹. Merleau-Ponty parle d'« habitude perceptive » et d'« habitude motrice » pour désigner la pratique régulière d'un monde qui, peu à peu, nous le rend « présent comme lieu familier de notre vie »² jusqu'à l'inclure dans notre sphère corporelle. À force de percevoir et d'agir dans un espace on finit par l'habiter. Merleau-Ponty décrit l'habitude comme une synthèse opérée par le corps :

S'habituer à un chapeau, à une automobile ou à un bâton, c'est s'installer en eux ou inversement, les faire participer à la voluminosité du corps propre. L'habitude exprime le pouvoir que nous avons de dilater notre être au monde, ou de changer d'existence en nous annexant de nouveaux instruments.³

L'habitude, c'est le corps qui comprend et qui apprivoise le monde avec lequel il est en relation. C'est ce processus qui est à l'œuvre dans la prise de vues : de fait, la caméra constitue un instrument qui prolonge l'expérience phénoménale de son opérateur, qui l'augmente. Lorsqu'il place son œil dans le viseur, l'objectif devient une extension de son regard tout comme le bâton de l'aveugle est une continuité de son toucher :

Quand le bâton devient un instrument familier, le monde des objets tactiles recule, il ne commence plus à l'épiderme de la main mais au bout du bâton. [...] Le bâton n'est plus un objet que l'aveugle percevrait, mais c'est un instrument avec lequel il perçoit. C'est un appendice d'un corps, une extension de la synthèse corporelle.⁴

¹ *Ibid.*

² *Ibid.*, p. 79.

³ *Ibid.*, p. 179.

⁴ *Ibid.*, p. 188.

La prise de vues est donc une perception instrumentée comme l'est la perception de l'aveugle par le biais de son bâton ; la caméra est incluse dans la sphère corporelle de son opérateur.

Et cette synthèse est encore augmentée par la temporalité de la prise de vues. Contrairement au peintre d'observation qui réalise sa toile sur le temps long — plusieurs heures, plusieurs jours ou plusieurs années — le caméraman vise le monde et l'enregistre en « temps réel », dans la temporalité même de son apparition. De ce fait, l'expérience de l'espace permise par la prise de vues est identique à celle que nous expérimentons chaque jour et dont Merleau-Ponty livre une description très proche de celle de Panofsky à propos de l'espace psychopshysologique. Le phénoménologue distingue deux modes d'existence du corps : le corps phénoménal ou corps propre, une sorte de corps *pour moi*, en intériorité ; et le corps objectif, mon corps pour autrui, en extériorité. Pour simplifier, le corps propre c'est le point de vue située, et le corps objectif c'est l'« œil unique et immobile », le point de vue de nulle part. L'un comme l'autre jouent un rôle important dans notre conscience de l'espace. Le corps phénoménal possède un espace premier, inné, qui n'existe que par lui : il est « à l'espace ». Le corps propre dispose alors d'une « spatialité de situation » et non d'une « spatialité de position » : elle se définit non par des coordonnées dans un espace mathématique mais par rapport à une action. C'est son intentionnalité, sa projection vers un but, qui donne au corps phénoménal un espace. De ce fait, il constitue toujours l'origine du repère, le point par rapport auquel tout objet détermine sa position :

Mon corps tout entier n'est pas pour moi un assemblage d'organes juxtaposés dans l'espace. Je le tiens dans une possession indivise et je connais la position de chacun de ses membres par un schéma corporel où ils sont tous enveloppés.¹

Dans le schéma corporel, décrit par Merleau-Ponty, l'espace mathématique ne s'applique pas car chaque organe constitue sa propre origine : les parties de mon corps « ne sont pas déployées les unes à côté des autres, mais enveloppées les unes dans les autres »². Ainsi, la position du corps propre est toujours absolue : mon bras est défini par sa position propre, connue de manière innée, et non par une position relative à un objet à côté de lui où par son inscription dans un repère spatial. Mais comment se constitue alors le corps objectif ? On a vu que le corps phénoménal est défini par son orientation vers un but : il est « à l'espace » dans une « spatialité de situation ». Mais c'est

¹ *Ibid.*, p. 127.

² *Ibid.*

justement cette intentionnalité qui entraîne, au moment de l'action, un retour du corps sur lui-même. Le corps phénoménal « se ramasse sur lui-même »¹ et soudain se voit en quelque sorte de l'extérieur, placé dans une « spatialité de position » : c'est le corps objectif. Celui-ci ne dispose plus d'un espace relatif dont il constituerait sans cesse l'origine, le point zéro ; au contraire, il est « dans l'espace », cet « espace infini, continu et homogène, [...] purement mathématique » qui était déjà celui de la perspective centrale. On le voit encore une fois, c'est à partir de l'espace corporel que l'espace objectif peut se déployer. C'est fondamentalement l'expérience corporelle qui permet la conception de l'espace rationnel et universel. Il serait impossible de définir un tel espace de manière abstraite sans avoir jamais vécu dans un corps auparavant : « il est essentiel à l'espace d'être toujours « déjà constitué » et nous ne le comprendrons jamais en nous retirant dans une perception sans monde. »².

La prise de vues hérite du corps de l'opérateur cette faculté de se mouvoir dans un espace et de comprendre la forme d'un objet en tournant autour de lui : c'est parce que je peux me déplacer que j'ai conscience des six faces du cube quand que je n'en vois toujours que trois. Alors même qu'on a voulu détacher la caméra de la subjectivité, c'est en fait ce point de vue incorporé qui fait toute la richesse de l'expérience la prise de vues. Contrairement aux autres régimes de figuration, le *faire image* de la prise de vues repose sur un espace en trois dimensions dans lequel l'opérateur peut « cadrer » en temps réel. Tout autre créateur d'image est forcé de se déplacer en esprit dans le monde imaginaire qu'il représente, seul le caméraman a la possibilité de se déplacer en corps. Surgit alors l'une des premières difficultés que devra affronter la prise de vues virtuelles. En effet, le rapport à l'espace que nous avons décrit semble aller à l'encontre de l'idée que nous nous faisons de l'expérience d'un monde virtuel numérique. À première vue, pas de corps dans une telle virtualité. Si tel était le cas, ce serait un coup important porté au concept de « prise de vues virtuelles », et sans doute la restriction de la prise de vues à la seule « prise de vues réelles ». Évitions cependant les jugements hâtifs : nous examinerons cette première difficulté dans la prochaine partie. Pour l'heure, il nous reste encore à introduire une seconde difficulté : le monde et son apparaître.

...Dans le monde

¹ *Ibid.*, p. 130.

² *Ibid.*, p. 292.

Puisque la prise de vues repose sur la perception, elle nécessite le point de vue d'un vivant mais aussi l'existence d'un monde comme horizon de cette perception. Ce monde dans lequel est « jeté » le vivant le précède et lui échappe toujours. C'est la métaphore du toucher qu'évoque Barbaras pour désigner la relation sensible : toujours en contact mais jamais coïncidence. En effet, le monde ne saurait être saisi, enveloppé comme un objet : c'est un apparaître. De fait, la relation sensible nous ouvre un monde plutôt qu'elle nous fait appréhender un *en soi* : « chaque expérience est une présentation de ce monde et non l'appréhension d'un objet »¹. Nous avons déjà évoqué le principe de la donation par esquisse : toute chose ne m'est donnée que partiellement à chaque instant et, si la succession de ses apparaîtres peut certainement enrichir l'expérience que j'en ai, je ne pourrai pour autant jamais prétendre atteindre un jour l'essence de cette chose en faisant la somme de ses apparitions. La question de l'apparaître soulève ici le paradoxe de l'identité dans la variété. Je perçois la table comme un seul et même objet alors même que mes perceptions ne font que varier. La table « s'esquisse dans une multiplicité de vécus » mais « chaque vécu est vécu de la même table ». Quel est ce mystère ? C'est que, selon la phénoménologie, le paraître se confond avec l'être : chaque apparition nous met en présence de l'être sans pour autant que nous ne puissions jamais le saisir. L'esquisse repousse toujours l'apparition de l'objet en lui-même. C'est en ce sens qu'elle est esquisse : elle préserve la transcendance de l'objet qui ne peut jamais être présent tout entier, en lui-même ; il s'oppose à moi. Ainsi, l'esquisse présente le même objet toujours changeant, au point qu'il « recule derrière sa propre présence »². Finalement, la chose perçue est toujours « inadéquate » et « contingente ». Descartes, dans son exemple fameux du morceau de cire, se sert du caractère changeant de la cire qui fond pour conclure à une tromperie de la perception. Nous percevons la cire dure et liquide comme deux choses différentes alors qu'il s'agit toujours de la même chose. De là, Descartes invite à soulever le voile de la sensation pour découvrir la chose en soi qui est pure identité, la substance de la cire toujours identique à elle-même. Mais on a vu le piège de l'intellectualisme cartésien qui met l'abstraction en première, ignorant l'expérience vécue à laquelle elle est pourtant toujours subordonnée. Contre l'exemple cartésien qui « rabat la perception sur la conception »³, Barbaras oppose la perception de l'eau. Il y a bien un sens à distinguer l'eau liquide de la glace ou de la vapeur. En réalité, c'est seulement pour la science que la matière se conserve ; pour nous le monde est pure changement.

¹ R. Barbaras, *La perception. Essai sur le sensible*, op. cit., p. 73.

² *Ibid.*, p. 41.

³ *Ibid.*, p. 22.

C'est donc dans un tel monde que la prise de vues opère. Encore une fois, ce qui la distingue des autres formes de figurations c'est qu'elle repose sur la perception : une présence concrète du monde qui est donné de manière sensible et non intellectuelle. Ce monde est alors une véritable force agissante capable de nous surprendre. Si le monde intelligible est le fruit d'une abstraction dont nous sommes la fondation, le monde sensible est à son tour la fondation de notre vécu. Le monde abstrait du peintre est dépendant de son esprit, le monde concret d'opérateur de caméra ne lui doit rien ; au contraire, il le précède et lui échappe. C'est plutôt l'opérateur qui lui doit une certaine reconnaissance. Cependant, l'autonomie du monde sensible ne doit pas nous faire conclure à une pure passivité du vivant qui l'occupe. Une telle conception reviendrait à une réinstauration de la séparation sujet-objet. Or c'est justement pour relier le percevant et le perçu que Barbaras requalifie le sujet en « vivant » et l'objet en « monde ». L'ontologie du Sensible implique d'aller à la rencontre du monde, à son contact : « Dans la mesure où la perception advient à même la chose, le sujet de la perception ne peut être qu'un sujet actif, qui se dépasse vraiment vers le monde, qui soit capable de le rejoindre en son lieu. »¹. Barbaras insiste sur la nécessité du mouvement dans la perception : je vois en déplaçant mon regard, mais je déplace mon regard en prévision de la vision. Et de la même manière que celle de la vision, l'expérience du toucher se fait dans le mouvement : la dureté s'éprouve avec la pression, la rugosité avec le frottement... Tout se passe comme si la sensation guidait le mouvement vers elle, avant même qu'il y ait une perception effective. C'est tout l'enjeu de la « spatialité de situation » décrite par Merleau-Ponty : le corps se perçoit d'abord par rapport à un but extérieur, donc comme un mouvement vers le monde avant d'être, par un retournement sur lui-même, une position dans ce monde. Il y a donc un aller-retour permanent entre l'« habitude perceptive » et l'« habitude motrice » au cœur de l'expérience vécue : le monde s'offre à la vue du photographe qui se met en mouvement pour aller à sa rencontre. La passivité de la perception s'accompagne toujours d'une activité : les cadrages sont des choix actifs, des actions vis-à-vis du monde dont l'agentivité augmente à mesure que l'opérateur de la prise de vues contribue à la mise en forme ce qu'il va capturer : construction du décors, placement des sujets, déclenchement de phénomènes sociaux, physiques... Ce ballet incessant de la perception et de l'action correspond précisément au processus de création artistique tel que le décrit Dewey. Indissociable de l'expérience du vivant dans son milieu, la production artistique repose sur une dynamique qui alterne sans cesse entre la perception, apparentée au domaine de l'« esthétique », et l'action, reliée à l'« artistique ». Au cours du processus créatif, l'artiste doit donc faire preuve d'une grande

¹ *Ibid.*, p. 58.

souplesse et manier constamment ces deux facultés qui s'alimentent de manière dialectique : l'esthétique procure le matériau que met en forme le sens artistique, mais ce sens artistique doit en permanence faire son autocritique en revenant toujours à l'esthétique. Comme la dynamique du corps propre et du corps objectif, l'artiste occupe à tour de rôle un point de vue interne et externe : regard vers son œuvre et regard sur son œuvre. C'est seulement au prix de cette harmonie de l'œil et de la main que la création artistique peut être fructueuse. En effet, Dewey met en garde l'artiste qui ne prête pas assez attention et qui s'abandonne à l'une des deux activités, délaissant l'autre : il risque alors d'agir de manière automatique et insincère ; regarder sans voir ou faire sans créer. Ainsi, une œuvre d'art doit pouvoir évoluer tout au long de sa production, sans cesse s'affiner sous les aller-retours de l'œil et de la main. Si ce processus créatif est valable pour tous les arts et toutes les figurations, son degré de liberté varie. Dewey cite l'exemple de l'architecture dont la lourde infrastructure des projets rend difficile des changements en cours de route. À l'autre bout du spectre, la prise de vues offre une très grande marge de manœuvre aux artistes qui savent s'en servir. C'est que le monde, on l'a vu, est loin d'être un objet figé ; au contraire il croît avec le vivant qui l'occupe. Donné par esquisses, il est toujours en mesure de surprendre celui qui enregistre ses apparitions. Les cinéastes qui cherchent véritablement à déployer tout le potentiel de la prise de vues savent garder un œil attentif et ouvert sur les phénomènes imprévisibles qui se produisent sous l'objectif, et même mieux, ils savent créer les conditions de la surprise. Le cinéaste Albert Serra raconte avoir tourné son film *Pacifiction*, doué d'une grande profondeur au sens phénoménologique — distance qui naît de l'approche du monde par le sentir —, avec un dispositif de trois caméras disposées en triangle de façon à perdre ses acteurs et actrices dans l'espace du plateau. De fait, l'acteur principal Benoît Magimel ne sait jamais lequel des trois angles sera conservé au montage. Il perd alors toute possibilité de jouer pour l'une des caméras : livré à une multitude de points de vue, il ne peut se voir de l'extérieur dans une spatialité de position et il ne lui reste plus qu'à s'abandonner entièrement à sa spatialité de situation, à n'être qu'une intériorité. Poussant le dispositif un cran plus loin, Serra va même jusqu'à dicter en temps réel son texte à Magimel qui ne le connaît pas à l'avance, provoquant une spontanéité parfaite de l'acteur. Ici, les comédiens comme les décors font partie du même monde façonné par le réalisateur espagnol pour favoriser la surprise au maximum. Serra sait, au moment où il écrit le scénario, qu'il ne s'agit que d'un brouillon de travail ; une grande partie du film surgissant au fur et à mesure du tournage de ce frottement dialectique entre la scène et le metteur en scène, l'un l'autre à tour de rôle surpris et surprenant. C'est là sans doute la principale distinction entre le cinéma et l'animation. Le réalisateur

d'animation qui se dit surpris par l'univers qu'il conçoit commet un abus de langage car en réalité c'est lui-même qu'il surprend. Le monde qu'il met en place est imaginaire et donc entièrement dépendant de son esprit. La surprise qu'il décrit est la même que celle de l'écrivain qui ne pense ses phrases qu'au moment où il les couche sur le papier. L'extériorisation des pensées peut certes produire un sentiment de révélation, mais c'est toujours une révélation de ses propres pensées : non pas que la surprise n'existe pas, mais c'est une surprise intellectuelle qui vient de l'intérieur. À l'inverse, le réalisateur qui s'appuie sur la prise de vues, qui plus est en lui donnant toute sa force comme le fait Albert Serra, est véritablement en présence d'une apparaître toujours extérieur et toujours nouveau. C'est une surprise sensible : c'est le monde, sans doute conditionné par la mise en scène mais néanmoins autonome, non-coïncidence avec le metteur en scène, qui prend celui-ci de court. Exploité à son plein potentiel, le tournage est, avant d'être la création d'un film, le développement d'une relation profonde et vitale avec le monde. Lorsqu'il travaille sur *Fitzcarraldo*, Werner Herzog passe plus de deux ans dans les profondeurs de l'Amazonie à mettre place, en particulier, la scène clé du grand bateau de métal blanc gravissant une colline boisée pour passer d'un bras à l'autre du fleuve. Si Herzog refuse catégoriquement d'avoir recours à une technique d'animation avec des maquettes en studio, c'est qu'il poursuit la volonté de faire advenir le monde de *Fitzcarraldo* ; il veut voir, vivre sensiblement la traversée de ce bateau :

Tel un chien fou qui s'est acharné sur une patte de chevreuil abattu et continue de secouer et de déchiqueter le gibier sans vie à tel point que le chasseur renonce à le calmer, une vision s'est emparée de moi: l'image d'un grand bateau à vapeur sur une montagne — le bateau sous la vapeur, utilisant sa propre force pour passer un versant pentu à travers la jungle, dans une nature qui anéantit les faibles comme les forts ; et la voix de Caruso, qui fait taire toutes les souffrances et tous les cris des animaux de la forêt vierge et arrête le chant des oiseaux.¹

Herzog lutte pendant deux ans contre, mais aussi avec, ce monde amazonien qui finit par imprégner le film de sa profondeur. À dix reprises il doit changer de lieu, d'acteurs, de bateau mais ce plongement total au sein de la forêt s'avère profondément créatif. Puisque le processus artistique requiert de l'artiste qu'il se projette dans son environnement, se frotte à lui, un environnement sans altérité et sans obstacle ne permettrait pas une expérience intéressante. La créature vivante ne pourrait prendre conscience d'elle-même sans se heurter à ce qui l'empêche, elle n'aurait rien à inventer, aucun moyen à mettre en œuvre pour parvenir à ses fins... Mais dans l'autre sens, un

¹ HERZOG Werner, *Conquête de l'inutile*, Nantes [Paris], Capricci Centre Pompidou, « Collection dirigée par Emmanuel Burdeau », 2009, p. 13.

environnement purement hostile ne serait pas constructif. Fitzcarraldo n'est pas passé loin de ce deuxième cas de figure, mais on sent bien au fond que le film tient du prétexte, là pour servir de justification à cette « conquête de l'inutile ».

Apparaît alors la seconde difficulté, sans doute la plus ardue, que devra surmonter la prise de vues virtuelles pour s'affirmer : fournir la preuve qu'elle est perception sensible dans un monde véritable, entier et vivant ; sans doute pas aussi épais que la jungle amazonienne, mais offrant au moins la possibilité d'une altérité constructive. C'est-à-dire que le monde numérique sous l'œil de la caméra virtuelle doit être « un monde dont la profondeur et l'indétermination répondent à l'insatisfaction qui caractérise la vie »¹. Le défi est important : pour l'heure, les mondes numériques soutenus par les moteurs jeu vidéo et autres dispositifs techniques de la sorte sont des simulations parcourues de « trous » et de « vides », de « crevasses ». Face à la complexité infinie du monde sensible, les mondes virtuels semblent condamnés à demeurer toujours des simulations partielles, des surfaces recouvrant des abîmes. À l'inverse du monde phénoménologique englobant le néant en son sein, et fournissant ainsi la possibilité d'un apparaître toujours nouveau, les simulations informatiques ressemblent plutôt au monde cartésien qui découle de la raison suffisante — un objet fini, entouré de néant. À la fin du documentaire précédemment évoqué *Knit's Island*, tourné dans le jeu *DayZ*, les réalisateurs montent une expédition avec une dizaine de joueurs qu'ils ont rencontré pendant le tournage virtuel, pour marcher jusqu'aux limites de la *map*, *i.e.* le territoire du jeu, et voir s'il y a la possibilité d'en sortir. Alors que la superficie totale du monde annoncée par les créateurs du jeu est d'environ deux cents kilomètres carrés, soit un carré de quinze kilomètres de côté, l'équipée marche plusieurs jours jusqu'à s'abîmer dans un *no man's land* qui finit par *bugger*. Perdus dans cet entremonde, la facticité de la simulation ne laisse aucun doute aux réalisateurs. Ses limites ainsi révélées mettent à jour l'illusion qu'elle entretenait : ce qui se faisait passer pour un monde n'en est pas un, c'est seulement un objet fini flottant dans le vide. Bien sûr, il ne s'agit pas de créer des mondes virtuels en tous points identiques au monde sensible pour la bonne raison que c'est impossible — il n'y a qu'un seul Monde dont font partie les simulations informatiques. En revanche, tout l'enjeu des mondes virtuels est de se rapprocher au plus près de l'expérience du monde vivant, c'est-à-dire de rendre possible de la manière la plus convaincante le déploiement du couple action-perception, de l'« habitude motrice » et de l'« habitude perceptive ». C'est de la possibilité d'« habiter » de tels mondes que dépendra le sort de la prise de vues virtuelles. À cheval

¹ R. Barbaras, *La perception. Essai sur le sensible*, *op. cit.*, p. 54.

entre la prise de vues réelles — *live action*, i.e. « action vivante » — et l'animation — action mécanique — elle peut basculer d'un côté comme de l'autre.

DEUXIÈME PARTIE. Tourner avec un ordinateur

IV. Expérience de la prise de vues réelles numérique

L'image-espace

Nous l'avons vu, ce qui fait la particularité de la prise de vues par rapport aux autres techniques de figuration, c'est qu'elle repose sur la perception. Dès lors, qu'elle soit argentique ou numérique, la prise de vues s'appuie sur un espace persistant et autonome par rapport à la caméra. Elle présuppose en cela un certain réalisme : l'existence extérieure, et non-dépendante d'un esprit, de la réalité perçue. Contre l'idéaliste qui se demande si la Lune existe encore lorsqu'il lui tourne le dos ou cesse d'y penser, l'opérateur de la prise de vues sait que le monde qu'il filme existe en dehors du cadre et que, lorsqu'il pivote la caméra, ce qui sort du champ ne sort pas de l'existence. C'est ce réalisme qui fonde, à la fois pour le cinéaste comme pour le spectateur, l'expérience spatiale du cinéma dont traite Antoine Gaudin dans son ouvrage *L'espace cinématographique*¹. Celui-ci part d'un constat : la question de l'espace est restée un angle mort des études sur le cinéma, relativement peu explorée par rapport à celle de la temporalité. Sans doute est-ce le cas parce qu'au moment où le cinéma est né, ce qui a marqué ses contemporains comme un élément vraiment novateur c'est cette illusion du mouvement si bien décrite par Bergson, l'espace ayant déjà été quant à lui investi à de nombreuses reprises par les arts précédents — nous pensons en premier lieu à l'entreprise picturale de la perspective. Pourtant, manquer l'espace serait manquer le cinéma comme art immersif : c'est sur l'expérience de l'espace vécu qu'il donne à vivre que repose, avant sa temporalité, la force sensorielle du septième art. Défendant l'importance de cet espace cinématographique, Gaudin propose le terme d'« image-espace » en référence à l'« image-temps » de Deleuze. Cette image-espace désigne, d'abord, l'espace vécu tel que le décrit la phénoménologie, et non sa représentation symbolique, ce qui permet alors de distinguer le cinéma des autres arts visuels :

L'espace au cinéma ne sera donc pas considéré comme un motif stable représenté par le film, et disponible pour la vue uniquement (dans la lignée d'une conception optique-systémique de l'espace

¹ GAUDIN Antoine, *L'espace cinématographique: Esthétique et dramaturgie*, Paris, Armand Colin, 2022. *Ibid.*

qui pourrait être commune à tous les arts visuels), mais comme un phénomène dynamique produit par le film, et engageant le corps du spectateur.¹

Au moins aussi importante que l'image-temps, l'image-espace consacre la rupture du cinéma vis-à-vis des arts qui le précèdent. La restitution de l'espace vécu fait de lui le premier véritable art immersif. Or, cette particularité repose sur la prise de vues. L'illusion du mouvement, produite par la succession rapide d'instantanés, est déjà présente dans l'animation dont c'est le principe même. En revanche, seule la prise de vues « capte et retranscrit de l'espace »². On a vu à quel point la caméra fait corps avec son opérateur pour permettre une perception instrumentée, augmentée. L'espace qu'elle enregistre, c'est l'espace du monde sensible qui existe hors de nous, pour notre perception. Si l'animation est capable de reproduire la sensation de l'espace vécu, c'est d'après les codes établis par le cinéma. Quand elle ne cherche pas à l'imiter, l'animation donne plus souvent lieu à des espaces flottants, étendues imaginaires de mondes sans corps. C'est que les réalisateurs de prise de vues réelles peuvent s'appuyer sur la persistance de l'espace physique pour construire la scénographie de l'espace cinématographique, c'est-à-dire organiser l'espace en fonction du champ et du hors champ. Dans le cinéma qui cherche à reproduire l'expérience de l'espace vécu, le hors champ existe et même, il est toujours présent. Comme dans le couple visible / invisible de Merleau-Ponty où le second terme constitue la structure du premier, le hors champ soutient le champ. D'autres arts visuels ont cherché, avant le cinéma, à rendre compte de l'espace vécu, à l'instar de la peinture de la Renaissance. Avec la perspective, il s'agissait en effet de donner l'impression d'un prolongement de l'espace hors du cadre : la composition du champ s'appuie sur le hors-champ. Mais la différence de l'opérateur de prise de vues avec le peintre du Quattrocento, c'est qu'il peut cadrer dans un espace persistant, non imaginaire ; et, surtout, il peut, avec le montage, agencer les plans dans le temps, donner à voir des vues différentes pour reconstruire une continuité spatiale. Citant Bazin, Gaudin souligne que l'espace du tableau est « centripète », « il ferme le tableau sur l'espace de sa propre composition » tandis que le cadre filmique est « centrifuge » : il nous invite à regarder à l'extérieur du cadre.

On comprend mieux l'importance du découpage qui fait passer le cinéma de la reproduction à l'expression. C'est que le réalisme spatial s'appuie sur le montage : la succession et la variation des plans permettent la construction de l'espace cinématographique. Or, c'est vraiment quand le cinéma s'approprie l'espace qu'il devient cinéma. Avant cela, il possède déjà le temps, mais

¹ A. Gaudin, *L'espace cinématographique*, op. cit., p. 11.

² *Ibid.*, p. 6.

l'expérience du tournage est statique, plate : c'est du théâtre filmé. L'image-espace naît de la domestication de l'espace par les cinéastes qui apprennent, d'abord, à bouger la caméra, et puis, à raccorder ensemble des plans aux échelles et aux angles de vue différents. Plus qu'une continuité temporelle, le découpage est surtout une continuité spatiale. Lev Manovich décrit ainsi le surgissement de l'espace cinématographique avec le passage du cinéma primitif au cinéma classique :

Le système de représentation du cinéma des débuts était de nature présenteielle : les acteurs jouaient pour le public, et le style était strictement frontal. La composition des plans mettait également l'accent sur la frontalité. En revanche, le cinéma hollywoodien classique place chaque spectateur à l'intérieur de l'espace fictif du récit. [...] Ainsi, l'espace ne sert plus de toile de fond théâtrale. Au contraire, grâce à de nouveaux principes de composition, à la mise en scène, à la conception des décors, à la cinématographie à grande profondeur de champ, à l'éclairage et aux mouvements de caméra, le spectateur est placé au point de vue optimal de chaque plan. Le spectateur est « présent » dans un espace qui n'existe pas réellement. [...] Grâce au montage, des images qui auraient pu être tournées dans des lieux géographiques différents ou à des moments différents créent l'illusion d'un espace et d'un temps contigus. Le montage est la technologie clé du XXe siècle pour créer des réalités fictives.¹

À partir de là, l'expérience du tournage gagne en spatialité pour ressembler de plus en plus à l'expérience phénoménologique de l'espace vécu. Contre le point de vue statique du théâtre filmé, le découpage libère la spatialité de la caméra qui devient un point de vue dynamique, pouvant se déplacer et, de ce fait, *habiter le monde*. Ce qui est en jeu, ici, c'est la restitution de la profondeur, cette distance irrattrapable entre le sujet percevant et le monde qui se laisse toucher sans jamais se laisser pénétrer. Le réalisateur qui se déplace et pense la continuité spatiale imprègne le film de cette profondeur sensible. À l'inverse, celui qui ne peut pas changer de point de vue reste à l'extérieur, condamné à n'enregistrer qu'une surface plane. Selon la phénoménologie, la profondeur caractérise notre expérience du monde. Le cinéma qui parvient à la restituer est donc doué d'un réalisme immersif d'autant plus fort. La profondeur est la clé de l'immersion. Gaudin y voit d'ailleurs le point fondamental sur lequel se joue l'opposition classique entre les frères lumières et

¹ L. Manovich, *The language of new media*, op. cit., p. 139-140., Traduction personnelle : « The early cinema's system of representation was presentational: actors played to the audience, and the style was strictly frontal. The composition of the shots also emphasized frontality. In contrast, classical Hollywood film positions each viewer inside the fictional space of the narrative. [...] Accordingly, the space no longer acts as a theatrical backdrop. Instead, through new compositional principles, staging, set design, deep focus cinematography, lighting and camera movement, the viewer is situated at the optimum viewpoint of each shot. The viewer is "present" inside a space which does not really exist. [...] Through editing, images that could have been shot in different geographic locations or in different times create an illusion of a contiguous space and time. Editing, or montage, is the key twentieth technology for creating fake realities. »

Méliès. Ce qui les différencie principalement, c'est leur rapport à l'espace : vues sur le monde et point de vue dynamique pour les frères Lumière¹ ; films-tableaux et point de vue statique pour Méliès. Cette opposition s'avère finalement être une opposition entre profondeur et planéité. Lorsque la caméra rejoint le point de vue mobile du corps, lorsque le découpage permet la constitution d'un espace cinématographique, le cinéma se dote d'une profondeur. Tourner un film à la manière des frères Lumière, c'est donc être dans l'espace du film, sentir l'espace que reconstitue ce film, être capable de toujours garder en tête la continuité spatiale tout en recherchant, à chaque plan, le meilleur point de vue possible sur l'action. C'est dans cette optique que se développe la grammaire du cinéma.

Cette grammaire, le septième art la doit d'abord à la prise de vues. C'est elle qui permet de concevoir l'écran comme un écran-fenêtre (une vue prise dans un monde qui l'englobe), tandis que l'animation dans sa forme classique conçoit plutôt l'écran comme écran-cadre. Le réalisateur de prise de vues, s'il peut concevoir la continuité spatiale a priori n'est pas tenu à cette priorité ; il peut tout autant — et c'est souvent le cas — réaliser la continuité spatiale a posteriori, au moment du montage. C'est là une grosse différence avec l'animation dont la continuité spatiale, si elle a lieu, doit toujours être conçue à priori. Et, au delà de la continuité entre les plans, c'est même chaque plan que l'animateur doit concevoir mentalement avant de le réaliser. Le cadreur, lui, les choisit plus qu'il ne les conçoit. Cette différence entre la prise de vues et l'animation est utile pour comprendre la distinction que fait Gaudin entre les réalisateurs qui s'intéressent à l'« espace global du monde » et ceux qui s'intéressent à l'« espace interne au cadre ». Les premiers sont engagés dans un rapport spatial au monde proche de l'expérience de l'espace vécu — cela correspond au paradigme de la prise de vues. Les seconds sont dans une approche plus plastique où l'écran est une surface à remplir qui constitue un espace symbolique — cette approche correspond au paradigme de l'animation. Mais cela ne signifie pas que tous les réalisateurs de la première catégorie soient des réalisateurs de prise de vues réelles, et inversement, que tous les réalisateurs du second groupe font de l'animation :

On retrouve ce principe encourageant le réalisateur à créer l'espace interne au cadre (tout y a sa place déterminée au sein d'une vision préalablement et totalement formée), plutôt que de le recueillir dans

¹ Il est à noter que le premier travelling est réalisé par un opérateur Lumière sur une gondole à Venise, les frères Lumière incitant leurs opérateurs à prendre des vues mobiles. Parmi les exemples fameux figure également le film de 1900, *Le Village de Namo, panorama pris depuis une chaise à porteurs*.

l'espace global du monde (avec l'idée de laisser le film ouvert aux fluctuations de ce dernier), chez des cinéastes comme Alfred Hitchcock ou Robert Bresson.¹

Hitchcock déclare en effet dans ses entretiens avec Truffaut qu'un réalisateur ne doit jamais « se laisser impressionner par l'espace qui se trouve devant la caméra », parce que le cinéma « permet d'obtenir tout ce que l'on désire, de réaliser toutes les images que l'on a prévues »². Hitchcock fait *storyboarder* l'entièreté de ses films dont le montage est dès lors déjà constitué avant même le tournage. Le « maître du suspense » demande à ses techniciens comme à ses acteurs d'exécuter à la lettre le plan pré-établi, au point qu'il déclare souvent n'éprouver aucun besoin de se rendre sur le plateau. De son côté, Bresson écrit dans ses *Notes sur le cinématographe* : « Comment se dissimuler que tout finit sur un rectangle de toile blanche ? Vois ton film comme une surface à couvrir »³. Ainsi, les cinéastes comme Bresson ou Hitchcock qui conçoivent l'espace de manière abstraite et a priori, et voient leur film « comme une surface à couvrir », quand bien même ils travaillent avec la technique de la prise de vues réelles sont en réalité plus proche du paradigme de l'animation. Au lieu de travailler avec un pinceau et une feuille de papier, Hitchcock se sert d'une caméra, mais il modèle le monde et compose son cadre avec le même degré de contrôle qu'un animateur : il transforme ses acteurs en marionnettes. À l'inverse, les réalisateurs d'animation qui s'engagent dans une relation dynamique et sensible avec le monde virtuel qu'ils mettent en scène, peuvent se rapprocher du paradigme de la prise de vues. L'expérience de la prise de vues réelles se différencie donc de l'animation avant tout par sa dimension a posteriori, là où l'animation est un a priori. C'est que la prise de vues réelles produit des images analogiques, toujours d'après un modèle donc toujours postérieures.

L'image analogique

Pour réaliser l'empreinte de son modèle, la prise de vues réelles, qu'elle soit argentique ou numérique, s'appuie sur un double phénomène physique : optique et électromagnétique. Concernant le premier, ce qui caractérise la prise de vues réelles c'est qu'elle constitue un dispositif optique produisant des images réelles d'objets réels. En physique, un objet réel désigne un objet qui est une

¹ A. Gaudin, *L'espace cinématographique*, op. cit., p. 44.

² HITCHCOCK Alfred et TRUFFAUT François, *Le cinéma selon Alfred Hitchcock*, Paris, Robert Laffont, 1966, p. 223.

³ BRESSON Robert, *Notes sur le cinématographe*, Paris, Gallimard, « Collection Folio Texte intégral », 1995, 137 p.

source de lumière à partir de laquelle les rayons lumineux divergent. À l'inverse, un objet virtuel est un objet qui donne l'impression d'être la source de la lumière alors que les rayons lumineux divergent en réalité d'un autre point, c'est le cas par exemple des objets dans un miroir. À partir de cette distinction entre objets réels et virtuels, la physique distingue à son tour images réelles et virtuelles. Une image réelle est l'image d'un objet réel pouvant être projetée sur un écran (écran de projection, pellicule, capteur numérique, rétine...) parce que les rayons lumineux qui divergent depuis la source lumineuse de l'objet réel passent par un dispositif optique (la lentille de l'objectif dans le cas de la prise de vue) qui les fait converger à nouveau vers cet écran. L'image réelle est donc l'image formée sur une surface par la convergence des rayons lumineux causée par un système optique. On parle dans ce cas de système optique objectif, catégorie à laquelle appartiennent les caméras argentiques et numériques. À l'inverse, une image virtuelle est l'image d'un objet réel ou virtuel causée par un dispositif optique qui fait diverger les rayons lumineux. Dès lors, l'image virtuelle ne peut être projetée sur un écran ; elle n'est visible qu'à travers le système optique qui la produit, par exemple une loupe. On parle dans ce cas d'instrument d'optique subjectif. D'après l'optique physique, il n'y a donc pas de différence entre la prise de vues réelles argentique et numérique. Toutes deux reposent sur un système optique identique produisant des images réelles : un objectif dont la lentille fait converger vers les rayons lumineux qui entrent dans le diaphragme vers un écran — la pellicule pour l'appareil argentique et le capteur pour l'appareil numérique.

Le second phénomène physique sur lequel s'appuie la prise de vues est la nature électromagnétique de la lumière, onde constituée de photons. Encore une fois, l'argentique et le numérique présente une identité de fonctionnement : après avoir traversé la lentille, les photons convergent vers une surface photosensible qui réagit à leur contact. L'argentique doit son nom aux cristaux d'argent utilisés pour la solution chimique photosensible qui recouvre la pellicule. En touchant cette surface, les photons modifient chimiquement certains cristaux. Une « image latente » se forme alors — l'empreinte chimique proportionnelle à l'intensité lumineuse. Pour révéler cette empreinte, il faut passer par l'étape du développement : on trempe la pellicule dans un bain révélateur qui retient les cristaux exposés et élimine les zones non exposées. Finalement, on obtient un négatif à partir duquel on peut tirer le positif. Les capteurs numérique, développés dans les années 1970 de façon à imiter parfaitement le fonctionnement de l'argentique, constituent une émulation de la pellicule. On distingue deux types de capteurs, CCD ou CMOS, au fonctionnement comparable : des millions de photodiodes (des composants semi-conducteurs sensibles à la lumière)

réparties sur une surface rectangulaire semblable à celle d'un photogramme¹. Chaque photodiode reçoit des photons et réagit en produisant une charge électrique proportionnelle à la quantité de lumière reçue. Ici, la réaction photochimique de la photographie argentique est remplacé de manière équivalente par l'effet photoélectrique : la production d'un signal électrique lors du contact entre un photon et une photodiode. Ce signal électrique correspond en quelque sorte à l'image latente de la photographie argentique. L'équivalent du développement, c'est alors la conversion numérique du signal. À cette étape, l'intensité électrique est quantifiée pour être convertie en valeurs numériques stockées dans un fichier binaire qui permettra ensuite d'afficher la photographie sur un écran sous forme d'une grille de *pixels*. Si la photographie numérique reste continue et indicielle jusqu'à l'étape du signal analogique, la numérisation de ce signal introduit la discrétisation de l'image. Mais, malgré cet espace de virtualité qui existe désormais avec le fichier binaire, entre l'« image latente » (le signal électrique) et l'image « développée » (la grille de pixels affichée à l'écran), la prise de vues réelles n'est pas remise en cause pour autant par le numérique. Elle s'appuie encore, avec cette nouvelle technologie, sur les mêmes phénomènes physiques de l'optique et de la nature électromagnétique de la lumière que l'argentique. Surtout, cette image numérique se forme toujours d'après un modèle. Ainsi, contrairement à un mauvais usage du terme « analogique » parfois employé pour désigner l'argentique, la prise de vues numérique est, elle aussi, d'ordre analogique.

Mais la définition physique reposant sur les phénomènes optiques et électromagnétiques ne semble pas suffisante pour qualifier précisément la prise de vues réelles. Si l'on s'en tient à ce critère, comment différencier alors le tournage avec des acteurs de chairs et d'os de la *stop-motion*, animation image par image de marionnettes, ou de la photographie des *celluloids*, surfaces transparentes sur lesquelles les animateurs peignent leurs *frames*. En effet, chacune de ces trois techniques repose sur l'enregistrement d'une réalité physique avec un appareil photographique, argentique ou numérique. Pourtant, seule la première correspond vraiment à de la prise de vues réelles, les deux autres appartenant au contraire à la catégorie de l'animation. Il nous faut alors préciser notre définition : la prise de vues réelle est l'enregistrement en temps réel d'un objet réel en mouvement. Les *celluloids* comme les figurines de *stop-motion* sont des objets réels au sens physique. Mais ils ne sont pas doués d'un mouvement « réel » dans le sens où l'est un système physique ou biologique. Leur mouvement dépend entièrement de l'animateur, ce sont des marionnettes. Le terme de « prise de vues réelles » est justement apparu pour se distinguer de

¹ Le « grand format » 24x36mm des appareils numériques haut-de-gamme est basé sur les dimensions de la pellicule 35mm

l'animation, y compris des techniques impliquant la prise de vues mais dont le mouvement n'est pas en temps réel. En effet, l'animation est caractérisée par l'image-par-image. La caméra de la *stop-motion*, par exemple, est en réalité plus un appareil photo qu'une caméra : elle ne filme pas à 24 images/seconde, mais plutôt elle prend une photo de chaque pose clé. Et il faut du temps à l'animateur pour modeler chaque pose, il peut donc s'écouler plusieurs minutes entre la prise de chaque photogramme. Ce nouveau critère qu'est l'enregistrement en temps réel d'un objet en mouvement vient s'ajouter à notre définition physique de la prise de vues réelles. Et il s'applique de la même manière à l'argentique et au numérique. Là encore, pas de différence notable entre les deux techniques.

Fort de ces précisions sur le fonctionnement de la prise de vues réelles et les critères qui permettent de la définir, nous sommes désormais en mesure de souligner la continuité de l'argentique et du numérique. Loin d'une rupture totale, d'une incommensurabilité, on remarque au contraire les ressemblances entre les deux techniques qui confinent parfois à l'identité, au moins fonctionnelle. C'est que la photographie numérique s'est développée d'après le modèle de la photographie argentique. Si le système optique reste inchangé, la principale différence entre les deux techniques réside dans le remplacement de la pellicule par le capteur. Mais là encore, l'analogie est flagrante : le capteur numérique est une émulation parfaite de la pellicule, et on peut en expliquer le fonctionnement avec les termes de l'argentique — contact des photons avec une surface photosensible, formation d'une image latente et développement. Sous cet angle, le numérique est une continuité parfaite de l'argentique. Et même, c'est un amplificateur de l'« effet cinéma », une amélioration technique comme le septième art en déjà connu avec l'arrivée du son ou de la couleur participant d'une plus grande immersivité. Finalement, tant que l'image issue de la prise de vues réelles numérique n'est pas retouchée, elle « ne ment pas ». Elle est du même degré d'objectivité que l'image argentique. On pourrait même dire que, tout comme elle, la photographie numérique est une empreinte. Cependant, un scrupule nous retient alors : certes la prise de vues réelles numérique, quand elle n'est pas retouchée, produit des images analogiques, d'après un modèle. Mais on a vu que le lien indiciel, qui court de l'objet réel au négatif de la photo argentique, est rompu pendant l'étape de la numérisation. La quantification du signal électrique analogique discrétise l'image jusqu'à là continue. Le « développement », qui correspond ici à l'étape de la numérisation, se fait, à l'inverse de l'argentique, sur un support immatériel : le fichier *.jpeg*, *.raw* ou n'importe quel autre format numérique qui transforme l'image un objet virtuel. Si la discontinuité physique qui surgit alors ne nous empêche pas, d'attribuer pleinement le statut de « prise de vues

réelles » au dispositif numérique, elle nous empêche cependant de parler d' « empreinte » au sens strict du terme. Pour être juste, il s'agit plutôt de la simulation d'une empreinte. La caméra numérique est un ordinateur dont le capteur simule la pellicule argentique. Et comme toute simulation, elle cherche à se rendre indiscernable de son modèle. Pourtant, la différence est toujours là, irréductible. Dans ce petit intervalle de discontinuité, dans ce vide entre le signal analogique et l'affichage pixellisé, se loge toute la différence entre le numérique et l'argentique. Ce néant est la promesse d'une libération.

La libération numérique

L'arrivée du numérique libère, en effet, la prise de vues réelles d'un grand nombre de contrainte physique, dans tournage comme dans l'exploitation de l'image après-coup. L'assouplissement des processus et des exigences techniques, en particulier la disparition du développement des négatifs en laboratoire entraîne un changement drastique dans l'expérience du tournage. Soudain, il est possible pour le réalisateur sur le plateau d'avoir un retour caméra en direct. Alors qu'avec l'argentique il fallait, dans le meilleur des cas pour les grosses productions qui en avaient les moyens logistiques, attendre au moins le lendemain pour voir le résultat des prises, le numérique permet un affichage vidéo en temps réel. Le réalisateur peut observer le jeu de ses acteurs, les mouvements de la caméra, le cadrage et la composition directement à travers l'œil de l'objectif. Les *combos*, petits écrans affichant le retour image et son de ce qui est filmé, fleurissent sur les plateaux, entraînant une redéfinition des rapports de force. En effet, le directeur de la photographie, rendu tout puissant par l'argentique puisqu'il était le seul, au moment du tournage, à *savoir* à quoi ressembleraient les images filmées, perd une part de son autorité avec le numérique. À l'époque où il fallait attendre au moins le lendemain pour avoir un aperçu du rendu colorimétrique, de la luminosité ou de la composition d'un plan, le réalisateur n'avait d'autre choix que de s'en remettre pleinement au directeur de la photographie. Avec le *combo*, le mystère disparaît : le réalisateur, mais aussi le producteur, les acteurs et n'importe quelle autre personne sur le plateau ont désormais accès à l'image en même temps que le chef opérateur, ce qui ouvre la voie à la possibilité, plus ou moins agréable pour celui-ci, de formuler des retours en direct. En parallèle de cette perte d'autorité, certains directeurs de la photographie regrettent également le mystère symbolique du développement. Il y avait, semble-t-il, quelque chose de rituel et d'excitant dans la découverte le lendemain du tournage des rushes développés pendant la nuit. Souvent, sur le plateau,

on organisait une petite projection qui rassemblait l'équipe. Cette excitation des *dailies*, même si elle a trouvé des remplaçants sous d'autres formes, a disparu telle quelle avec le numérique. Cette nostalgie rejoint la critique plus générale d'une perte de matérialité avec l'abandon de l'argentique. Les artefacts de la pellicule, les rayures et le grain, mais aussi les limitations techniques comme la durée assez courte des bobines, sont accueillies, par les adeptes de l'argentique, comme des contraintes légitimes, propre à l'existence matérielle du film, et produisant de l'engagement. Dans cette appréhension, le chef opérateur est un artisan qui, à l'instar du sculpteur ou de l'ébéniste, travaille la matière directement au contact du monde. La dimension indicielle de l'image argentique et l'étape du développement, au-delà de l'autorité qu'elles peuvent lui apporter, donnent surtout un sens au travail du directeur de la photographie. Face à cet ancrage physique, l'immatérialité du numérique peut être ressentie comme désengageante. Et ce sentiment n'est pas réservé aux chefs opérateurs, des réalisateurs comme Quentin Tarantino refusent de tourner en numérique pour des convictions du même ordre :

J'aime la cinématographie qui permet d'obtenir le résultat souhaité le jour même. [Dans le cinéma], on peut faire un tirage décoloré et toutes sortes de petites choses. Mais si ce n'est pas capturé le jour même, alors on ne l'a pas capturé. [...] Avec le numérique, oubliez ça. On peut littéralement faire tout ce qu'on veut. [...] Dans un monde où tout est possible, rien n'a de sens.¹

Mais cette vision du numérique produisant une forme de désengagement nihiliste — tout est permis alors plus rien n'a de sens — n'est pas partagée par tout le monde, de majoritaire elle est même devenue minoritaire. Des réalisateurs comme Georges Lucas, dont le film *Star Wars, épisode I : La menace fantôme* marque un tournant pour la prise de vues numérique, embrassent au contraire cette technologie avec un grand enthousiasme. C'est qu'elle apporte avec elle une liberté qui peut, contrairement à l'avis de Tarantino, produire un engagement avec le monde plus important encore que l'argentique.

En effet, si le numérique est rapidement adopté par les grosses productions hollywoodiennes en raison de la flexibilité qu'il permet pour les retouches des effets spéciaux, il est aussi utilisé par des réalisateurs indépendants dans une approche assez différente : se rapprocher du monde filmé dans une sobriété de moyens et d'artifices. Rompant avec l'appareillage contraignant et imposant du

¹ Tiré d'une interview sur *Youtube* rapportée par le site *No Film School* : <https://nofilmschool.com/why-tarantino-not-fan-digital-cinematography>. Traduction personnelle : « I like cinematography that you get what you do on the day. [In film] you can do a bleached print and there are all kinds of little things you can do. But if it's not captured on the day, then you didn't capture it. [...] When it's digital, forget about it. You can do literally anything you want. [...] In a world where you can do anything, nothing means anything. »

tournage argentique qui met une distance entre le réalisateur et le monde, le numérique permet au contraire un rapprochement. Cet allègement de l'infrastructure du tournage se joue à plusieurs niveaux. D'abord, les caméras numériques offrent des solutions plus compactes et plus légères que les lourdes caméras argentiques qui doivent embarquer avec elles des magasins de pellicule. Certes, les caméras utilisées sur le tournage de blockbusters comme *Star Wars*, puisqu'elles cherchent à imiter le rendu visuel de l'argentique sont en réalité tout aussi imposantes. Mais pour les réalisateurs prêts à embrasser l'esthétique *trash* des premières caméras DV, s'ouvre désormais la possibilité de filmer caméra au point. Certains films marquant de la fin des années 90 portent la marque de la caméra DV, parmi lesquels *Festen*, premier film du Dogme95, dont les images semblent provenir directement de l'enregistrement amateur d'une véritable fête de famille. Et en plus de leur légèreté, les nouvelles caméras numériques sont presque dix fois plus sensibles à la lumière que la pellicule ce qui assouplit les contraintes lumineuses du tournage — les lourds *spots* ne sont plus nécessaires. Et la sensibilité ISO, fixée et limitée par le modèle de pellicule, devient ajustable selon la nécessité de telle ou telle scène. Enfin, la question stockage disparaît quasiment avec le numérique. Il devient possible de tourner un film en un seul plan séquence, là où les bobines de pellicule limitaient la longueur des prises — la capacité de stockage standard du magasin d'une caméra 35 mm permettant de stocker 1000 pieds de film soit environ 11 minutes à 24 images par seconde. Les réalisateurs peuvent désormais tourner virtuellement autant qu'ils veulent sans être limités par la longueur de la bobine, et, surtout, sans se heurter à des limites financières. En effet, la pellicule coûte cher, à l'achat mais aussi au développement. À l'ère argentique chaque mise en marche de la caméra est à prendre au sérieux, et certains contrats entre les producteurs et les réalisateurs stipulent le droit de ces derniers à un nombre limité de prises. Le numérique constitue ainsi, à tous ces égards, une véritable libération qui, contrairement à l'idée reçue d'une technologie tournée vers la surenchère d'effets spéciaux, permet également l'essor de mouvements cinématographiques indépendants revendiquant un retour à la simplicité et une économie de moyens. Le cinéma avait déjà connu de telles libérations avec des progrès techniques qui avaient permis l'avènement de mouvements semblables. La mise au point des caméras portables 16mm et d'une pellicule plus sensible à la lumière avait déjà rendu possible, en son temps, le *Néoréalisme* et puis la *Nouvelle Vague* dont les réalisateurs s'étaient aventurés hors des studios pour tourner au plus près du réel. En ce sens, l'arrivée des caméras DV ressemble à ce mouvement libérateur qui accompagna les caméras légères prisées par les cinéastes des *Cahiers du Cinéma*, ouvrant un nouveau chapitre sur le thème classique du *retour au réel* dans l'histoire du cinéma. Il est par

exemple révélateur que *Festen*, le premier film du *Dogme95*, mouvement initié par Lars Von Trier et Thomas Vinterberg sur la base d'un « vœu de chasteté », ait été tourné avec une caméra *DV*. Alors même que la 9^e règle du manifeste précise que tous les films du mouvement devront être tournés en 35 mm, gage d'authenticité et de simplicité, il est très intéressant de relever cette entorse au règlement. C'est que dans sa recherche d'épure et de vérité, le *DV* est un choix plus évident que le pourtant très simple 35mm. Le *Dogme95*, encore imprégné de la pellicule, annonce déjà l'ère du cinéma numérique indépendant. Le *Mumblecore* qui se développe au début des années 2000 aux États-Unis se saisit des nouveaux formats digitaux comme la promesse d'un cinéma sans autorisation, sans producteur et sans argent. Certains réalisateurs poussent la logique de la caméra *DV* jusqu'au bout en adoptant les smartphones comme des caméras : le réalisateur Sean Baker, figure du cinéma indépendant américain, tourne intégralement son long-métrage *Tangerine* à l'*iPhone 5s*. Difficile de faire mieux en termes de légèreté. En plus du faible budget, Baker explique que le dispositif lui a permis de tourner en ville sans avoir à interrompre la circulation ou à bloquer des rues, permettant une très grande flexibilité créative et une meilleure mise en condition des acteurs non-professionnels. C'est ici la démonstration que la légèreté des moyens et la réduction au strict minimum des contraintes techniques n'est nécessairement synonyme d'un engagement moins fort. La souplesse du tournage de *Tangerine* semble plutôt se rapprocher du processus de création artistique que nous avons décrit dans le chapitre précédent : un élan constant fait d'aller-retours entre le monde et l'artiste, et, plus le médium permet une flexibilité importante, plus les conditions de créations sont idéales. Le numérique offre aux réalisateurs la possibilité d'un engagement plus dynamique avec un véritable tournage a posteriori, nécessitant une préparation moins lourde en amont, pour une véritable rencontre avec le monde.

Le poids que gagne le tournage, et sa corollaire, le montage, avec le numérique apparaît de manière flagrante dans les statistiques des *shooting ratios* — le ratio des images filmées par rapport à la durée du montage finale. Avec le numérique, le ratio explose. Pendant l'âge d'or d'Hollywood, des années 30 aux années 60, le *shooting ratio* était d'environ 10 pour 1 soit 25h de *rushes* pour un film de 90 minutes. Mais certains réalisateurs avaient un ratio encore inférieur. Celui d'Hitchcock était par exemple de 3:1, ce qui n'est pas surprenant de la part d'un réalisateur préparant minutieusement son tournage à l'avance. Il est intéressant de noter que ce ratio de 3:1 est très proche de celui des films d'animation qui oscille souvent entre 1:1 et 2:1¹, ce qui va dans le sens de

¹Il est tout aussi intéressant de noter que le *shooting ratio* de l'animation n'a pas augmenté avec le passage au numérique, la production des images animées étant toujours, voire plus difficile qu'avant.

notre catégorisation, volontairement provocante, d'Hitchcock comme réalisateur d'animation. En effet, un ratio de tournage très faible implique une préparation a priori d'autant plus importante, tandis qu'un ratio élevé déplace la quantité de travail en aval du tournage, au moment du montage. Cette étape est une autre des caractéristiques fondamentales de la prise de vues réelles — le « montage » de l'animation correspondant plutôt à l'étape du storyboard — et elle gagne en importance à mesure qu'augmente le poids du tournage. On doit d'ailleurs à Georges Lucas, pionnier de la prise de vues numérique, les premières machines de montage numérique mises-au-point au sein de son studio d'effets spéciaux *ILM*. Désormais, c'est le montage aussi qui est libéré par le numérique. S'il fallait y réfléchir à deux fois avant de trancher la pellicule, il est enfin possible de couper à volonté pour un nombre virtuellement infini d'itérations. Le numérique augmente ainsi le couple tournage-montage, paradigmatique de la prise de vues réelles, modèle inverse de celui du storyboard-animation. Cet accent sur la création a posteriori séduit des réalisateurs comme Albert Serra qui cherchent à retranscrire des expériences ignorées à l'avance plutôt qu'à « mettre en boîte » des histoires déjà constituées. Pour reprendre l'exemple de son film *Pacifiction*, le tournage a donné 540 heures de rushes pour un film de 2h45, soit un *shooting ratio* de quasiment 200:1. Avec un tel écart, il est clair que la réalisation du film est d'abord l'expérience d'une découverte, voire d'une révélation :

Quand je fabrique un film, même pendant le montage, je n'ai aucune idée de la signification de ce que je suis en train de faire. Je ne fais jamais d'essais, le scénario je suis le seul à le connaître, je ne le donne jamais aux acteurs. Tout mon système est basé sur la non-communication avec les acteurs et les techniciens. C'est une performance où rien ne se répète, il y a une fatalité dans chaque moment d'enregistrement et dans le montage, c'est la même logique. Il y a des choses que j'aime pour des raisons très différentes et avec le monteur on essaie de créer un film cohérent uniquement avec les moments que j'aime, ce qui fait que je n'ai aucune idée de ce que le film va signifier : c'est comme un processus scientifique. Mon film porte un discours, mais je ne sais pas lequel, c'est mystérieux.¹

Une telle expérience de tournage est rendue possible avec le numérique, ce qu'affirme lui-même Serra qui insiste régulièrement sur la dimension technique de ses projets. À cet égard, le numérique permet une augmentation de la prise de vues, et même, une augmentation de l'espace cinématographique.

¹ SERRA Albert, « Par les temps qui courent ». <https://www.radiofrance.fr/franceculture/podcasts/par-les-temps-qui-courent/albert-serra-sur-les-tournages-j-aime-qu-il-y-ait-une-rebellion-8696386>

L'augmentation de l'espace

La caméra-ordinateur ouvre la voie à toutes les manipulations possibles. Une partie des cinéastes adoptent le numérique comme un moyen de réduire l'écart entre eux et le monde, de se rapprocher au plus près de leur sujet, faisant de la révélation une méthode de tournage ; d'autres, à l'instar, de Georges Lucas s'emparent de la technologie d'abord pour la possibilité qu'elle offre de dépasser l'espace physique. En parallèle du cinéma indépendant, le cinéma d'effets spéciaux constitue également un pionnier de l'utilisation numérique. En effet, le stockage immatériel des images sous forme de fichiers informatiques permet un aller-retour bien plus efficace avec les studios d'effets spéciaux ; surtout, cette nouvelle forme facilite grandement l'intégration invisible des « images réelles » et des images de synthèse. Plus besoin de scanner la pellicule pour la numériser au risque de perdre en qualité, il est possible de travailler directement à partir du fichier enregistré pendant le tournage : le flux est direct entre la caméra du chef opérateur et l'ordinateur de l'infographiste. Se démocratise alors le fameux « fond vert », ou « fond bleu », qui permet des *set extensions* — l'augmentation de l'espace physique du plateau. L'incrustation, qui plonge des acteurs de chair et d'os dans des décors de synthèse d'une manière indiscernable, devient l'effet spécial de loin le plus populaire et le plus emblématique. Cela fait particulièrement sens pour des films comme la saga des *Star Wars* qui se déroulent dans des univers fantastiques où les décors terrestres ne sont pas suffisants. Soudain, on film les acteurs devant des fonds verts ou bleus et les voilà arpentant la croûte glacée d'une planète lointaine. Bien sûr, le procédé de l'incrustation n'est pas né avec l'informatique ; il est en réalité aussi vieux que le cinéma : Méliès l'utilisait déjà, Disney avait mis au point l'« écran jaune » pour son Magicien d'oz de 1939 — un procédé chimique qui permettait, les acteurs jouant devant une teinte particulière de jaune, de les isoler parfaitement du fond. Mais c'est le numérique qui en démocratise vraiment l'usage, faisant du tournage sur fond vert une norme du film à effets spéciaux, et pas seulement. C'est qu'avec l'informatique naît le principe du *compositing*, « le processus consistant à combiner plusieurs séquences d'images animées et éventuellement des images fixes en une seule séquence à l'aide de logiciels de composition spéciaux »¹. Il s'agit en quelque sorte d'un montage, non pas des images les unes à la suite des autres, mais au sein même de l'image, un photomontage. Manovich voit dans le *compositing* un élément clé des nouveaux médias : ce montage dans l'image s'accorde parfaitement avec la logique de sélection et

¹ L. Manovich, *The language of new media*, op. cit., p. 130., Traduction personnelle : « the process of combining a number of moving image sequences and possibly stills into a single sequence with the help of special compositing softwares »

de modularité que mettent en place les outils numériques : « Le compositing numérique illustre bien un principe général de la culture informatique : assembler plusieurs éléments pour créer un seul objet homogène »¹. Les logiciels de *compositing* comme *Photoshop*, *After Effects* ou encore *Nuke* fleurissent et augmentent drastiquement les possibilités du photomontage, jusqu'à là limité par la matérialité de la pellicule. S'il était possible de gratter le visage de Trotsky pour le retirer d'une photographie officielle, ou de superposer plusieurs images les unes sur les autres, les contraintes matérielles se faisaient vite ressentir. Avec le numérique, le seul obstacle devient la puissance de calcul qui permet dans tous les cas une augmentation exponentielle du nombre de couches et d'éléments juxtaposées : l'un des plans d'*Avatar II* a nécessité plus de 10 000 *nodes*, i.e. couches de retouches, dans le logiciel *Nuke*. À titre de comparaison, une retouche invisible sur une photographie physique peut difficilement excéder les cinq ou dix couches. Mais est-il vraiment juste de parler de « montage » dans le cas du *compositing* ? Manovich défend, à juste titre, qu'on peut y voir, au contraire, les signes d'un anti-montage dont le but est justement de gommer la juxtaposition, la coupe, pour donner l'illusion d'une continuité parfaite. En cela, le *compositing* est en quelque sorte l'opposé du collage, qui, en tant qu'il affirme les coupes, correspond mieux au principe du montage. D'un autre côté, Manovich voit déjà dans la peinture le principe du « montage dans l'image » pour simuler la réalité. Et le cinéma classique invente à son tour le montage temporel. Finalement, le cinéma d'effets spéciaux numériques combine la peinture et le cinéma classique en synthétisant les deux types de montage. Le *compositing* rejoint ainsi le montage temporel dans la création d'une continuité spatiale, constitutive de l'espace cinématographique, qu'il augmente significativement. *Star Wars I* qui sort en 1999 est à 95% constitué de séquences composées de manière numérique ; il met en scène des galaxies dont l'immensité pouvait seule être contenue par le virtuel. Les plateaux de tournage de tels films offrent alors un curieux spectacle : de grands hangars verts où les murs et le sol, recouverts de la même couleur uniforme, se confondent et nous font perdre la notion de l'espace. Les quelques acteurs face à la caméra jouent devant un fond vert en guise de décors, interagissent avec des boules vertes perchées sur de longues tiges qui seront plus tard remplacés par des paysages et créatures fantastiques. Il peut être difficile pour les acteurs de se repérer dans un tel espace et il faut sans doute à Emilia Clark, interprète de Daenerys Targaryen dans *Game of Thrones*, fournir un très grand effort d'imagination pour arriver à voir, dans la boule verte qui s'agite devant elle, la tête énorme d'un de ses dragons.

¹ *Ibid.*, p. 132., Traduction personnelle : « Digital compositing exemplifies a more general operation of computer culture: assembling together a number of elements to create a single seamless object. »

De nouveau le risque du désengagement surgit. Le *compositing* prend beaucoup de temps, des dizaines voire centaines d'heures, au risque de perdre la spontanéité du tournage. Il faut parfois attendre des semaines ou des mois pour voir le résultat d'une prise, entre temps, le réalisateur, et surtout les acteurs qui n'ont pas forcément accès à des images de prévisualisation, n'ont d'autres ressources que leur imagination. Par ailleurs, les effets spéciaux comme les *set extensions* impliquent une lourde infrastructure qui rigidifie le processus créatif. En particulier, ils nécessitent une forte préparation en amont du tournage : il faut prévoir les effets qu'on produira, et, en conséquence, les actions précisent des acteurs et des actrices qui perdront dès lors en marge de manœuvre. Surtout, pour s'assurer de ne pas entraver les possibilités de retouche en post-production, les tournages se font plus neutres. Il s'agit d'enregistrer un maximum d'informations non contraignantes de manière à pouvoir tout modifier en aval du tournage :

Avec le virtuel, je dis bien virtuel, et non pas numérique, il ne faut surtout pas se laisser imprégner d'une atmosphère lors de la prise de vues, surtout pas la renforcer, nous dit-on. Tout se fera après, et il vaut mieux que le document original soit neutre pour permettre le plus de manipulations numériques possibles.¹

Les retouches numériques peuvent donc entraîner une forme de glissement de la prise de décision créative, qui passe du tournage à la post-production. Le chef opérateur Charles Van Damme pointe alors le risque d'un

désinvestissement sur le plateau où le travail du directeur de la photographie se limite fondamentalement à rassembler des éléments qui ne prendront leur forme définitive que beaucoup plus tard. C'est après le tournage, devant l'écran de l'ordinateur que se prennent les décisions de fond, à froid, alors qu'en travaillant "à l'ancienne", les décisions se prennent à chaud, sur le plateau, et elles engagent le directeur de la photo et le metteur en scène d'une manière irréversible.²

Comme l'illustre la formule ironique « Fix it in post ! » qui courent sur les plateaux, certains réalisateurs et producteurs se laissent en effet gagner par une tendance à repousser les choix à plus tard, quitte à créer des tâches fastidieuses pour les infographistes qui s'efforcent de régler péniblement des problèmes qui auraient plus l'être simplement au moment du tournage. Mais surtout, la neutralité, qui sert de sécurité en garantissant la plus grande marge de manœuvre en post-

¹ VAN DAMME Charles, « Le numérique, le virtuel et l'effet cinéma », in . https://www.unionchefsopérateurs.com/wp-content/uploads/2020/01/charlie_vandamme_numerique_virtuel.pdf

² *Ibid.*

production, produit effectivement un engagement créatif plus faible sur le plateau. Comment, par exemple, tourner dans un fond vert qui, par son uniformité, son absence d'éléments saillants, est un non-espace dans lequel le déplacement et le changement de point de vue n'ont plus aucun sens ?

L'hybridation de la prise de vues

Le désengagement provoqué par le *compositing* est surtout dû au décalage qu'il entraîne, reléguant la prise de décision créative du tournage à la post-production. En effet, le *compositing* requiert un temps considérable, auquel s'ajoute le temps de rendu des images composées qui peut prendre jusqu'à plusieurs mois. Mais, contre ce décalage temporel, l'informatique graphique développe depuis une trentaine d'années une technologie qui n'a de cesse de progresser et qui apparaît comme la solution sur-mesure : les moteurs de rendu en temps réel, ou *realtime*. Nous avons déjà évoqué la petite révolution que constitue le *realtime* au début des années 90, d'abord dans le jeu vidéo auquel il apporte des graphismes en trois-dimensions, et puis très rapidement dans le domaine du cinéma. Les moteurs de rendu « classiques » utilisés par l'industrie des effets spéciaux calculent, de manière très précise, le parcours des rayons de lumière dans les scènes virtuels, produisant des images photo réalistes ; mais cela a un coût : le temps de rendu peut atteindre plusieurs jours pour une seule image. À raison de 24 images par seconde, on imagine la quantité phénoménale de calculs que nécessite un long-métrage dont la quasi-totalité des plans recourent au *compositing* d'images de synthèse. À l'inverse de ces moteurs de rendu dits « *offlines* », le temps réel s'appuie sur des algorithmes de calcul, certes moins réalistes, mais bien plus rapides : les images sont calculées dans l'instant qui précède leur affichage à l'écran. C'est que le temps réel, d'abord mis au point pour le jeu vidéo, est conçu en vue de l'interactivité. Les images de synthèses de l'industrie du cinéma doivent être très réalistes, en revanche leur temps de rendu peut être assez long puisqu'elles ne sont calculées qu'une fois pour toute, de manière prédéterminée. En revanche, les images du jeu vidéo se doivent de répondre instantanément aux injonctions des joueurs, le réalisme passant en second. Pourtant, les graphismes assez rudimentaires des premiers moteurs *realtime* n'empêche l'industrie des effets spéciaux de s'y intéresser très tôt, justement dans l'optique de rendre plus flexible les pratiques du tournage. Dès le début des années 2000, des outils s'appuyant sur la technologie du temps-réel sont développés et utilisés sur les tournages, donnant alors naissance à la « virtual cinematography ».

Parmi les premiers dispositifs déployés, on peut citer la *motion capture* en temps réel : la retranscription instantanée, sur une créature de synthèse, des mouvements d'un acteur portant une

combinaison de capteurs. Sur *Le Seigneur des anneaux*, le monstrueux Gollum est rendu en temps réel sur les moniteurs du plateau d'après les mouvements de son interprète, Andy Serkis. Ce *compositing* en direct permet de visualiser au moment même du tournage le rendu final des scènes, facilitant le travail du réalisateur, du chef opérateur et des acteurs qui savent désormais à la rencontre de quel monstre va leur personnage. Même si, dans les premiers temps du *compositing realtime*, l'image est encore assez grossière, elle offre un aperçu déjà conséquent. L'opérateur qui regarde le moniteur de la caméra voit Gollum à la place son acteur qui s'agite, bardé de capteurs, au milieu du plateau : le numérique poursuit et augmente, avec le *realtime*, la perception instrumentée initiée par l'argentique. La caméra permet désormais de percevoir directement dans un monde virtuel, sans passer par l'étape, longue et discontinue, de la post-production : la prise de vues réelles s'hybride, elle devient prise de vues dans deux mondes à la fois. Et ce n'est que le début. La *motion capture* ne concerne pas seulement les acteurs, elle peut aussi être employée pour capturer les mouvements de la caméra. Ainsi, lorsque l'opérateur bouge dans le monde physique, la caméra virtuelle, qui sert à calculer l'image de synthèse compositée avec l'image réelle, se meut de la même manière, garantissant une incrustation parfaite, *seamless*. La *virtual camera*, contrôlée par les mouvements de la caméra physique, offre une solution bien plus convaincante que la méthode souvent laborieuse du *tracking* sur l'image rasterisée. Surtout, la *motion capture* de la caméra, combinée au *realtime* permet un engagement physique de l'opérateur. Celui-ci peut disposer directement, dans son retour-caméra, d'une prévisualisation du *compositing* mélangeant l'image réelle et ses composantes virtuelles. Les mouvements de la caméra physique pourront ensuite être récupérés par l'infographiste qui fera de manière propre et définitive le *composing*. À la prise de vues, s'ajoute donc, avec les procédés de *motion capture*, la prise de mouvement. Les déplacements de Gollum comme ceux de la caméra virtuelle ne sont pas le fait d'un animateur, derrière son ordinateur, qui déplace des objets virtuels avec sa souris, mais le résultat de mouvements réels dans le monde physique. Et si, dans un premier temps, les moteurs de rendu en temps-réels étaient utilisés uniquement pour leur interactivité à des fins de prévisualisation, la qualité de leurs graphismes a drastiquement augmenté dans les dernières années, au point que leurs images ont commencé à être utilisées de manières définitives, comme alternative aux images produites par les moteurs *offlines*. Ce progrès qualitatif a donné lieu, en particulier, au développement de la *virtual production* : le tournage, non plus devant un fond vert, mais devant un « mur LED », un grand écran, souvent en arc de cercle de façon à entourer au maximum le plateau, parfois couvrant également le plafond, qui affiche, en temps réel, un décor virtuel en fonction des

mouvements et des angles de la caméra physique de façon à respecter l'effet de parallaxe. S'appuyant sur la *motion capture*, la caméra virtuelle « filme » dans les mondes virtuels en respectant la même direction, la position, et même l'optique, de la caméra physique, garantissant une parfaite continuité entre le décor physique, sur le plateau, et virtuel, « derrière » l'écran. Bien sûr, de l'extérieur l'illusion est visible, mais dès lors qu'on regarde à travers la caméra, dans son axe exact, l'illusion devient indiscernable : le monde virtuel affiché à l'écran est une continuité parfaite du monde physique. Ici, il ne s'agit plus de *compositing* dans le sens où l'image est déjà-là, entière ; il n'y a pas d'assemblage ou de photomontage : tout ce qui apparaît dans le retour-caméra n'est pas le fait d'un traitement logiciel mais existe physiquement. Le *compositing* est déjà réalisé dans le monde extérieur. La *virtual production* constitue ainsi une solution bienvenue pour l'équipe de tournage. De nouveau, les acteurs peuvent voir dans quel environnement leurs personnages sont censés être plongés. Le directeur de la photo peut ajuster en direct les lumières du plateau et du monde virtuel pour s'assurer de leur bonne continuité — ce qui n'était faisable, jusqu'à là, qu'en post-production. L'opérateur caméra peut ajuster son cadrage en fonction du décor : de nouveau bouger la caméra a un sens. Le réalisateur peut entrer dans une relation dynamique avec le monde, avec une liberté créative d'autant plus forte qu'il peut demander des changements de décor en direct. Dans les moteurs *realtime* utilisés comme *Unreal Engine* ou *Unity*, il est aisé de déplacer une montagne, de faire pivoter un paysage à 180 degrés ou encore de changer l'orientation du soleil. Même si le tournage avec un *LED wall* implique toujours une grosse infrastructure et une lourde préparation, celle-ci est censée permettre une plus grande flexibilité le jour du tournage. Surtout, cet immense écran augmente, virtuellement à l'infini, et de manière sensible, l'espace du plateau.

Finalement l'irruption de toutes ces techniques de visualisation en temps réel sur le tournage permettent de réintroduire de l'engagement et d'augmenter l'espace du tournage. Si le plateau est toujours limité physiquement, l'expérience spatiale est drastiquement agrandie par les mondes virtuels. L'être-au-monde du réalisateur est augmenté et parfois même curieusement dédoublé. Une forme étrange de *virtual production* abandonne complètement la prise de vues pour n'être qu'une prise de mouvements ; elle consiste à « tourner » avec une fausse caméra équipée de capteurs de mouvements et des acteurs également revêtus de capteurs. Aucune images « réelle » n'est enregistrée pendant ce « tournage » qui a pourtant bien lieu physiquement. Les opérateurs caméras courent après les acteurs pour les cadrer, mais, en guise de caméra ils n'ont qu'un écran accroché à un *steadycam* dont les mouvements sont retranscrits sur une caméra virtuelle. Le retour-caméra

affiché sur leurs écrans, et qui leur permet de cadrer, correspond en réalité à ce que « voit » la caméra virtuelle dans l'espace 3D rendu en temps-réel et dans lequel les mouvements des acteurs physiques sont également retranscrits sur des créatures virtuelles. Les cadresurs et les acteurs sont donc dans deux mondes à la fois. Tout en évoluant dans l'espace physique, il leur faut prendre en compte cet autre espace parallèle, l'espace virtuel, qui est là comme un fantôme. C'est une pure virtualité (au sens mental, non-numérique) pour les acteurs qui n'ont que leur imaginaire pour le concevoir ; et une virtualité numérique pour les *cameramen* et tous ceux qui ont un retour vidéo du moteur *realtime*. La caméra physiques n'est plus ici que symbolique : nulle caméra avec une optique et un capteur photosensible, seulement un *steadycam* monté d'un écran dont les mouvements sont captés en temps réel. La prise de vues est ici parfaite hybridation, elle a déjà un bon pied dans le virtuel. Des deux éléments qui définissent la prise de vues réelles (capture d'images réelles selon la définition physique et capture d'objets en mouvement en temps-réel), il n'en reste plus qu'un seul : la capture du mouvement ; pour ce qui est de l'image, elle est virtuelle.

Prise de volume : un point de vue de toutes parts

À travers les différentes techniques d'augmentation de l'espace de cinématographique, que ce soit via le rapprochement du réalisateur avec le monde grâce à la légèreté des caméras numériques ou par le biais d'une perception instrumentée qui donne à percevoir des mondes virtuels directement superposés au monde physique, l'unité et la singularité du point de vue reste immuable. C'est que la prise de vues, on s'en souvient, hérite du caractère toujours déjà situé de l'être-au-monde. La phénoménologie pose le point de vue incorporé comme condition de toute expérience. La prise de vues n'échappe à ce prérequis : elle est toujours vue depuis un point situé dans le monde : l'« œil » de la caméra aligné sur celui de l'opérateur. Mais la phénoménologie de la perception décrit également comment la mobilité du corps répond à cette irréductibilité du point de vue : à chaque instant je ne vois que d'un seul endroit, mais j'ai la capacité de me déplacer, de changer mon point de vue. Perception instrumentée, la prise de vues épouse l'expérience perceptive du corps : elle est toujours, à chaque instant, point de vue unique dans le monde mais elle peut se déplacer et varier. Les mouvements de caméra et le montage ont été inventés justement pour pallier cette singularité du point de vue qui ne peut voir que d'un endroit à la fois. Le montage permet de se déplacer dans l'espace cinématographique, de varier les points de vue, participant en cela à l'immersivité du cinéma. Dès qu'il cesse d'être du théâtre filmé, le septième art rejoint l'expérience

corporelle pour devenir immersif. Quoi qu'il en soit, la caméra, pas plus que le corps, n'ont le don d'ubiquité : ils peuvent se déplacer mais ne peuvent jamais être dans deux endroits à la fois. Même la prise de vues hybride que nous venons de mentionner, si elle s'appuie sur deux caméras (une caméra physique et son double virtuel), ne forme toujours qu'un seul et unique point de vue. La caméra physique, prolongation du point de vue de son opérateur, et elle même prolongée par la caméra virtuelle. Dans la chaîne qui va de l'opérateur physique à la caméra virtuelle l'unité du point de vue est toujours préservée.

Les choses changent complètement avec la « prise de volume ». Nous utilisons ici ce terme pour nous référer à différentes techniques comme la photogrammétrie ou le *gaussian splatting* qui permettent de combiner une multiplicité de vues pour reconstituer un volume. Avec ces techniques une succession d'images — ou d'apparaîtres du monde — sont combinées entre elles pour constituer un objet ou une scène en trois dimensions. Cette technique est souvent utilisée par l'industrie du cinéma et du jeu vidéo pour créer des doublures virtuelles d'objets physiques, d'acteurs ou de décors sur lesquelles on pourra ensuite réaliser toutes les manipulations possibles dans un moteur de rendu 3D. Pour réaliser le scan d'un objet, on peut distinguer deux méthodes différentes. Soit une seule caméra qui prend une succession de clichés ou une vidéo en tournant autour de l'objet à scanner ; soit une multitude de caméras disposées sur une armature en forme d'igloo au centre de laquelle est positionné l'objet physique de manière à le photographier simultanément sous tous les angles possibles. Dans la première méthode, qui n'implique qu'une seule caméra, le scan s'effectue dans le temps. La règle de la singularité du point de vue est ainsi toujours préservée au sens strict : on enregistre toujours une succession d'images, une seule image à la fois. Mais au lieu d'être mises bout à bout sur une timeline, ces images sont montées en parallèle, comme si elles avaient toutes été prises en même temps. Elles sont combinées par un logiciel qui parvient, en les recoupant entre elles, à déterminer la position de chaque point de vue de chaque image dans un espace en 3D pour, ensuite, reconstituer le volume exact de l'objet et le recouvrir d'une texture qui est le fruit de la combinaison de toutes les photos. Cette méthode correspond finalement au point de vue mobile décrit par la phénoménologie : j'appréhende l'espace et la voluminosité des objets, les six faces du cube par exemple, en tournant autour d'eux. De la même manière, pour scanner un objet avec une seule caméra je dois tourner autour de lui pour couvrir toutes ses faces. Et de ce fait, ce mode de scan est fragilisé par sa temporalité. Comme le point de vue subjectif qui ne peut saisir un objet qu'un côté à la fois, le scan 3D, avec une seule caméra, n'a

accès qu'à des esquisses du monde toujours en mouvement, succession d'apparaître jamais coïncidant avec lui-même. Le temps assez court que nécessite un scan (l'intervalle entre la première et la dernière photographie de la série allant de quelques minutes à quelques heures) permet sans difficulté de scanner des objets statiques avec une seule caméra. En revanche, pour un objet mouvant comme les rameaux d'un arbre sous le vent ou, bien sûr, un être humain dont il est impossible d'obtenir l'immobilité parfaite, le scan 3D inclura forcément des flous de bouger comme une photographie à longue exposition. En fin de compte, l'agencement de la séquence d'image sur un axe vertical, simultanée, et non horizontal, successif, ne permet pas d'échapper à la marque du temps ni de s'affranchir de la fragilité de la perception subjective qui n'a toujours accès au monde que par la succession de ses apparaître, un seul à la fois.

La seconde méthode, celle qui consiste à prendre la multitude des vues nécessaires à la reconstruction de l'objet scanné en 3D au même moment, semble pallier cette fragilité du point de vue singulier. Avec suffisamment de caméras pour couvrir toutes les faces de l'objet à scanner, nul besoin de temps ni de déplacement : tous les points de vue peuvent être capturés au même instant. La première méthode correspondait à la perception mobile du corps propre, la seconde n'est pas sans rappeler le point de vue objectif, de toutes parts, auquel cherchait justement à répondre la phénoménologie. Force est de constater que la simultanéité des photographies apporte à cette seconde méthode une robustesse que n'avait pas la première. C'est la raison pour laquelle cette technique est utilisée principalement pour réaliser des doublures numériques d'acteurs. Ils s'installent au milieu de la cage d'appareils photographiques disposés de manière à couvrir toute la surface de leur corps, tous ses côtés à la fois. On leur demande alors de prendre des poses, des expressions faciales différentes, dont on capture de véritables instantanés en 3D — vision absolue de l'objet saisi de toutes parts. On peut alors les importer dans les moteurs de rendu 3D pour les exposer à toutes les retouches, les illuminer, les animer, les intégrer à des univers numériques et, finalement, en produire des images, capturées, cette fois, par une caméra virtuelle. En réalité, cette méthode de scan est loin du point de vue absolu. C'est le point de vue de plusieurs endroits à la fois mais ce n'est pas le point de vue infini de tous les endroits, le point de vue absolu étant reconstitué virtuellement par le logiciel qui interpole les images des appareils photographiques physiques. Quoi qu'il en soit, ce mode de prise de vues est une photographie en trois dimensions, qui rompt avec la singularité du point de vue. La prise de volume s'appuie sur l'ubiquité des points de vue.

La seconde méthode de scan, point de vue de toutes parts, ouvre alors la voie à la prise de vues en quatre dimensions, c'est-à-dire les trois dimensions de l'espace auxquelles s'ajoute celle du temps. La prise de vues réelles gagne donc ici une dimension : elle en comportait déjà trois, avec les deux dimensions de l'espace plat et le temps, désormais elle enregistre, non plus des images, mais des volumes en mouvement. Ce procédé du scan 4D est très lourd avec la photogrammétrie : il faut prendre des photographies de tous les angles simultanément, reconstituer et, surtout, afficher le modèle en 3D vingt-quatre fois par seconde, ce qui demande une puissance de calcul trop importante pour la plupart des ordinateurs. Cette technique est utilisée lorsqu'une grosse précision est requise, notamment pour des effets spéciaux photoréalistes, mais elle est utilisée avec parcimonie et ponctualité. Impossible d'imaginer un long-métrage, ou même un court-métrage, tourné en 4D de cette manière. En revanche, le *gaussian splatting* — technique bien plus légère que la photogrammétrie puisqu'au lieu de reconstituer une géométrie faite de surfaces virtuelles, elle positionne des points de couleur dans l'espace — permet un affichage en temps réel dans un navigateur internet. Cette technique a d'ailleurs connu un récent bond technologique avec l'emploi du *machine learning* pour l'interpolations entre les points de vue, diminuant le nombre de caméras nécessaires. Au moment où nous écrivons, le start-up chinoise *4DV.ai* vient de mettre au point une algorithme particulièrement performant pour la capture et la reconstitution de vidéos en quatre dimensions dont des démonstrations sont disponibles sur son site internet¹. Certains algorithmes de *machine learning* permettent même de générer des objets 3D à partir d'une seule image en « devinant » les faces dissimulées. Il devient possible de concevoir des tournages en quatre dimensions, que ce soit avec plusieurs caméras tournant en simultané ou même une seule caméra couplée à du machine learning.

Mais si le *gaussian splatting* 4D permet un enregistrement en volume, le mode de visualisation est lui toujours subordonné au point de vue unique. En fin de compte, lorsqu'elle est affichée sur un écran, une vidéo en 4D ne diffère pas en apparence d'une vidéo classique : c'est toujours une succession d'images, un seul point de vue à la fois. Elle possède, en revanche, une particularité qui la distingue : il est possible de se déplacer *dans la vidéo*, de choisir un point de vue particulier au moment du visionnage. Il s'agit ici d'une sorte de vidéo 360° améliorée, capable d'enregistrer non seulement des décors statiques en volume mais aussi des acteurs en mouvement. Deux solutions d'exploitation s'offrent alors pour le cinéma. Soit on décide d'abandonner au spectateur le choix du point de vue de la caméra : il est alors promu au rang de cadreur.

¹ <https://www.4dv.ai>

Techniquement, cette possibilité est permise par différents supports : le classique écran 2D couplé à un dispositif interactif comme une simple souris permettant d’orienter l’angle de la caméra et de se déplacer dans l’image ; ou bien, de manière plus novatrice, le casque de réalité virtuelle — l’utilisateur se servant désormais de son corps pour contrôler la caméra virtuelle dans une vidéo pourtant tournée entièrement avec des caméras physiques. Soit, le réalisateur, ou le chef opérateur, conservent la prérogative du cadrage. Rien ne change pour le spectateur qui reste passif, soumis au point de vue unique du film dont la séquence d’images qui le compose est fixe et non interactive. En revanche, pour le réalisateur c’est une manière de reléguer le choix du point de vue à plus tard, après le tournage. Nous avons déjà vu que le numérique pouvait entraîner le report d’un grand nombre de choix de mise en scène du tournage à la post-production. Avec la prise de volume, c’est désormais le cadrage qui peut être remis à plus tard. Il suffit alors d’importer le fichier vidéo 4D dans un moteur de rendu en temps réel pour le « relighter » — changer sa mise en lumière en ajoutant des éclairages virtuels — et, surtout, pour enfin choisir le point de vue unique et définitif de chaque plan du film. Celui-ci est du ressort de la caméra virtuelle qui prend le relais des caméras physiques du plateau, et à partir de laquelle est déterminé le rendu du moteur 3D. Une telle méthode dédoublerait ainsi le tournage : d’abord enregistrement de la matière première (espace physique, acteurs de chair et d’os, etc...) dans le monde, et puis mise en scène plus classique, cette fois dans le moteur de rendu 3D — choix d’un point de vue unique et d’un découpage avec les cadrages de la caméra virtuelle. La partie virtuelle du processus constituerait ainsi le moment des choix artistiques concernant le découpage, alors que la première phase du tournage physique concernerait plutôt les problématiques de scénographie et de directions des acteurs. Cela permettrait une très grande flexibilité dans les mouvements de caméras (possibilité de faire accomplir à la caméra des mouvements impossibles physiquement) et dans le montage. Le réalisateur pourrait tourner et assembler en parallèle : rendre les plans au fur et à mesure de ses besoins de montage. Un tel procédé constituerait, après les différentes variantes déjà évoquées de la *virtual production*, une autre forme d’hybridation parfaite entre prise de vues réelles et prise de vues virtuelles. Difficile de pousser plus loin cette hybridation sans basculer entièrement dans la pure prise de vues virtuelles, abandonnant la persistance de l’espace physique pour la semi-persistance des mondes numériques.

V. Expérience de la synthèse 3D

La synthèse d'images : entre animation et génération

On parle de « synthèse d'images » pour désigner le processus informatique de génération de nouvelles images d'après d'un modèle ou d'une description mathématiques. Cette « génération » est, plus précisément, un calcul : calcul de l'image pour son affichage sur un écran. Le terme d'« image de synthèse » désigne toutes les images entièrement générées par ordinateur. Il ne s'applique donc correctement qu'aux images vraiment nouvelles : le *compositing* ne constitue pas à proprement parler une opération de synthèse puisqu'il assemble des morceaux d'images (photographiques ou de synthèse) déjà existants. Il travaille directement avec des *pixels* — des *picture elements* — qu'il sélectionne, manipule, « colle » et superpose, donnant lieu à une nouvelle configuration. Si de cette configuration résulte une image certes différente, nouvelle en ce sens, le *compositing* n'a pas permis pour autant l'émergence d'éléments visuels fondamentalement nouveaux, qui n'existaient pas avant l'opération. De même, contre une idée répandue, la retouche photographique, ne constitue pas plus une forme de synthèse d'image car dans ce cas l'image originale a été capturée par un dispositif optique. Il s'agit, là encore, d'un cas de *compositing*.

Quels exemples pourrions-nous alors fournir de synthèses d'images véritables ? Il en existe une grande variété qu'on pourrait répartir en plusieurs catégories. D'abord, la peinture numérique, qui constitue l'une des formes de synthèse les plus anciennes. Nous avons déjà évoqué le cas de la première interface graphique, *Sketchpad* développée par Ivan Sutherland : un logiciel permettant de dessiner directement sur l'écran de l'ordinateur avec un stylet. Il s'agit bien là d'une image de synthèse, i.e. entièrement générée de manière informatique. La peinture numérique se répand rapidement et les tablettes graphiques investissent les studios de dessin animé, remplaçant les tables d'animation et les tables lumineuses. Si ces bouleversements provoquent, dans les premiers temps, des oppositions de la part de certains animateurs attachés au papier et au crayon, les avantages conséquents de la peinture numérique finissent par emporter l'adhésion du plus grand nombre. Aujourd'hui, rares sont les studios d'animation 2D travaillant encore sur papier. Autre signe de la popularité, des peintres prestigieux comme David Hockney revendiquent même leur usage de l'*iPad* ou de l'*iPhone* dans leurs productions, certaines des expositions d'Hockney étant entièrement constituées d'œuvres numériques. C'est que la peinture numérique permet un grand confort d'utilisation : des logiciels comme *Photoshop* ou *Procreate* offrent une immense quantité de

« brushes » qui simulent toutes sortes de pinceaux et de techniques (du crayon à papier à la peinture à l'huile en passant par les pastels ou l'aréographe...), et de grande possibilité de *compositing* et de retouches, additionnant les calques et leurs modes de fusion... Ce dernier point, en particulier, est un argument de taille pour l'animation qui devait, jusque-là, composer avec les limites physiques de techniques comme le *celluloid*. Mais la peinture numérique, si elle fait bien partie de la synthèse d'images, en constitue un cas périphérique, du moins pas le plus intéressant pour rendre compte du processus. En effet, elle a de la technique un usage assez limité : ici, nulle simulation d'après un modèle mathématique ; la marge du logiciel est réduite à son minimum, il se borne à retranscrire sous forme de *pixels*, sans doute en ajoutant quelques effets de *brush*, la pression directe sur l'écran du stylet manipulé par l'utilisateur. Le degré d'automatisation est donc minime.

À l'autre bout du spectre, le procédé emblématique de la synthèse d'images est ce qu'on appelle communément la *3D* : la génération d'images à partir d'une scène en trois dimensions baignée dans une lumière virtuelle. La 3D, dont on a déjà retracé l'histoire, constitue sans doute le meilleur exemple de synthèse d'image puisqu'elle met en œuvre un haut niveau de simulation, notamment à travers le calcul de la lumière virtuelle d'après les lois physiques. Ici, le degré d'automatisme est très élevé, et le terme de « génération » semble plus juste, là où il apparaissait légèrement galvaudé dans le cas de la peinture numérique. Contrairement au logiciel de peinture qui ne génère pas grand-chose, hormis quelques effets de textures, se bornant plutôt à une simple interprétation ou même traduction des gestes directs de l'utilisateur, le moteur de rendu 3D se voit au contraire déléguer une part conséquente de la création de l'image. La marge du logiciel est donc très importante ; il constitue, dès lors, un véritable intermédiaire dans la création : il n'y aurait aucun moyen de générer une image de synthèse 3D sans moteur de rendu, là où Hockney n'a pas un besoin essentiel de son *iPad* pour peindre ses paysages forestiers qu'il pourrait tout aussi bien produire, d'une manière quasi-identique, avec une toile et un pinceau. Le logiciel 3D est déterminant, ce qui montre son degré d'autonomie et sa dimension proprement générative.

Aux deux grands exemples de la peinture numérique et de la 3D, s'ajoutent tout un tas de processus de synthèse d'images à des fins artistiques, plus ou moins autonomes et plus ou moins génératifs. Parmi eux, on pourrait citer plusieurs exemples intéressants. D'abord, l'animation de « marionnettes 2D » dans des logiciels comme *After Effects* ou *Flash*, qui consiste, non pas à dessiner, mais à déplacer des formes géométriques relativement simples dans un espace en deux

dimensions, à la manière de l'animation de papier découpé. Le réalisateur Will Anderson a produit de beaux films avec cette technique, notamment son court métrage *Betty* récompensé aux *Baftas* qui utilise de manière convaincante les possibilités numériques pour transformer son *canvas* 2D en un véritable espace. Ensuite, l'art génératif ou le *creative coding* dont le principe consiste, comme leurs noms l'indiquent, à produire des contenus visuels évolutifs en fonction d'un programme. Le logiciel *Touch designer*, très utilisé en particulier pour des performances en direct, permet de générer des visuels à partir d'algorithmes qui fonctionnent souvent comme des interpréteurs de données. L'interface du logiciel permet de le connecter à toutes sortes d'*inputs* dont on configure l'influence sur la génération d'images. Dans l'œuvre *Coded Nature* du studio d'artistes *Drift*, un capteur de mouvement permet, par exemple, aux spectateurs de contrôler avec leur corps les mouvements d'un nuage de particules sur un écran géant. Autre outil prisé des *creative coders*, le langage de programmation *p5.js* basé sur *java script* a été justement développé à des fins artistiques par une petite communauté très active et enthousiaste. En codant toutes sortes d'opérations mathématiques (allant du simple *noise* au *machine learning*), *p5.js* permet de générer des œuvres dynamiques en 2D comme en 3D. On peut avoir un bon aperçu des possibilités créatives du langage en se rendant sur son site internet¹.

Toutes les techniques citées, lorsqu'elles incluent une dimension temporelle, sont alors communément rangées dans la catégorie de l'animation. L'adoption, en particulier, des techniques 2D (peinture numérique, animation de marionnettes) et 3D par les studios d'animation ont contribué à une ouverture des techniques d'animation, jusque-là cantonnées aux techniques traditionnelles du dessin et de la *stop-motion*, sans toutefois entraîner de véritable redéfinition de l'animation elle-même. Celle-ci est encore et toujours, à l'ère numérique, l'art de donner vie, par la mise en mouvement, à des formes, des dessins ou des objets inertes. L'animation, nous l'avons déjà évoqué, sous-entend une forme de création continuée, une création en direct de l'animateur qui insuffle la vie à ses créatures quasiment image par image. En cela, la peinture numérique est sans ambiguïté du côté de l'animation. Son mode de création direct, qui repose sur une très faible automatisation, constitue bien l'une des techniques sur lesquelles l'animateur peut s'appuyer pour créer le mouvement pose par pose, instant par instant. En effet, malgré les nombreux outils intégrés dans les programmes de peinture numérique pour faciliter la tâche des animateurs, le principe est toujours le même que celui l'animation 2D traditionnelle : il faut dessiner chaque image à la main. Aucun des logiciels standards de l'industrie, comme *TVPaint* ou *ToonBoom*, n'est en mesure de générer des

¹ <https://p5js.org/sketches/>

interpolations entre deux dessins. On comprend alors le bouleversement qu'annonce l'intelligence artificielle générative capable, justement, de générer de telles interpolations. Mais, pour l'heure les *workflows* d'animation 2D n'ont pas vraiment intégré l'IA et l'image par image reste la norme.

Qu'en est-il en revanche d'autres modes de synthèse d'image comme la 3D ou le *creative coding* ? La classification de telles techniques dans la catégorie de l'animation apparaît d'emblée plus équivoque. Toutes deux reposent sur une très forte dimension générative : elles donnent lieu à des images calculées d'après des modèles mathématiques, permettant ainsi une grande configuration et évolutivité de l'image. Surtout, ce sont des cas de création par l'intermédiaire d'un logiciel. À l'inverse du processus d'animation 2D par peinture numérique dans lequel l'animateur agit directement sur l'image finale (il la dessine), le *creative coder* ou l'infographiste 3D ne travaillent pas sur l'image rasterisée elle-même, mais sur un objet virtuel (le code ou la scène 3D) dont le logiciel se servira seulement ensuite pour générer l'image de synthèse. C'est donc une création indirecte qui laisse au logiciel une marge de manœuvre dans laquelle se situe sa dimension générative. Bien sûr, parler d'animation à propos de films 3D comme ceux de *Pixar* répond à une certaine logique : il y a bien une animation, quasiment image par image, de personnages 3D par un animateur qui intervient en direct sur sa création. Mais c'est là une seule étape, certes emblématique bien que minoritaire, dans un processus qui en compte des dizaines et qui reposent pour la plupart sur un mode de création indirect. Et l'animateur 3D s'appuie lui-même beaucoup sur les interpolations calculées par le logiciel entre ses « poses clés ». L'automatisation apparaît donc comme un premier élément venant brouiller les pistes. Le principe de l'animation comme création continuée du mouvement, image par image, est déjà moins évident avec des techniques de synthèse d'image intégrant un fort degré d'automatisation, et donc de génération. L'animateur n'intervient plus sur le mouvement de chaque image mais seulement sur certaines poses clés dont les poses intermédiaires sont calculées automatiquement par l'ordinateur. Parfois, il n'y a même plus d'animateur comme dans les cas de la *motion capture* ou de procédés purement génératif capable de produire des séquences d'images en continu sans intervention direct, ce que permet par exemple *Touch designer*. Et, au delà du mouvement, l'animateur n'agit pas non plus, avec la 3D comme le *creative coding*, sur chaque image ; il n'agit même sur aucune image puisque, on vient de le voir, il ne travaille pas en direct comme le dessinateur sur sa toile, mais sur des objets virtuels qui donneront, à leur tour, au moment du rendu par le logiciel, des images. La « génération » des images synthèse — la véritable génération, automatisée, qui délègue une grande part de la

production de l'image au logiciel — apparaît donc comme une première remise en cause de la notion d'animation au profit de la prise de vues virtuelles.

Le « rendu », moment clé de la synthèse d'images, désigne le procédé de calcul qui permet de passer d'une description mathématique de l'image virtuelle à l'image rasterisée, affichable sur un écran. En cela, le rendu semble correspondre à une forme de prise de vues virtuelles : prise de vues dans un espace virtuel. En effet, les moteurs de synthèse génèrent leurs image d'après une réalité préexistante. Tout comme la photographie est l'enregistrement d'un *ça a été*, le *ça a été* est ici un *ça a été* virtuel. D'ailleurs, le rendu ne concerne pas au même niveau tous les types de synthèse d'images : on ne parle pas de « rendu » pour la peinture numérique car, le rendu au sens propre sur lequel elle s'appuie est une rasterisation en direct, d'après le toucher du stylet manié par l'utilisateur et non d'après une description mathématique. À l'inverse, le terme est omniprésent à propos de la 3D, qui en constitue l'exemple paradigmatique : le calcul d'image à partir d'un espace tridimensionnel dans lequel sont simulés des phénomènes physiques. Dans ce type de rendu, au sens fort du terme, rendu d'après une réalité mathématique, toutes les images avant d'être des *raster images* sont d'abord des images vectorielles, c'est-à-dire des descriptions mathématiques d'un espace. La position de chaque point dans cet espace virtuel est indiquée sous la forme d'un vecteur avec une direction et une origine précise dans un repère. Il est ainsi possible de décrire la forme et le mouvement de n'importe quel objet virtuel à partir des points qui le composent. Le repère vectoriel peut être de plusieurs dimensions : une, deux ou trois pour les exemples déjà cités de synthèse d'image¹. Le rendu est donc le processus de transformation de l'image vectorielle en image rasterisée, i.e. le passage d'une image abstraite qui n'existe que comme description mathématique à l'image concrète, visible, de la grille de pixels. Les images vectorielles sont donc véritablement virtuelles, images latentes dont le potentiel est actualisé par le rendu. De leur côté, les images rasterisées sont virtuelles seulement au sens numérique du terme : elles sont affichées sur un écran et sont stockée dans un fichier binaire sous la forme d'une suite de chiffres codants la couleur de chaque *pixel*. Le procédé de rendu, s'il consiste toujours à retranscrire l'espace mathématique de l'image vectorielle dans l'espace en deux dimensions de l'écran, diffère en revanche selon le nombre de dimensions du repère. Le rendu à partir d'un espace 1D concerne la visualisation de courbes ou de *noise* à une dimension. Il met en œuvre une interpolation entre les valeurs pour permettre leur représentation en deux dimensions. Le rendu à partir d'un espace 2D est le plus

¹ La génération d'image par intelligence artificielle, que fera l'objet du prochain chapitre, s'appuie, elle, sur des espaces probabilistes qui peuvent atteindre plusieurs millions de dimensions

évident : l'espace de l'écran étant lui-même en deux dimensions, il s'agit simplement d'attribuer à chaque point de l'espace vectoriel le *pixel* qui lui correspond. C'est le cas de l'animation de marionnettes en 2D avec des logiciels comme *Flash*. Enfin, le rendu 3D implique d'abord une projection de l'espace 3D pour obtenir une image en deux dimensions, puis l'attribution d'une couleur à chaque pixel en fonction d'un calcul qui combine les propriétés physiques des objets dans la scène 3D, leurs textures et leur illumination. Cette double particularité d'un espace en trois dimensions et d'une simulation de phénomènes physiques, à commencer par la lumière, marque la spécificité du rendu 3D, *physically based*, par rapport aux autres modes de rendu. C'est cette différence qui fait de la 3D le cas emblématique de la synthèse d'images, au point qu'on les confond souvent¹. Par sa proximité avec l'espace physique, le rendu 3D constitue bien l'archétype de la « vraie » synthèse d'images qui se fait à partir d'un modèle mathématique, à partir d'une réalité virtuelle calculée, voire capturée, pendant l'étape du rendu. C'est ce type de rendu qui nous intéresse pour la prise de vues virtuelles.

Le rendu 3D : *physically based* et espace semi-persistant

Le rendu 3D *physically based* est un processus complexe qui entretient des liens étroits avec la prise de vues réelles. Comme elle, il se fait d'après un espace tridimensionnel dans lequel l'opérateur peut se déplacer. Surtout, il repose sur deux éléments essentiels de la prise de vues : la caméra et la lumière. La chercheuse en informatique graphique Marie-Paule Cani résume ainsi la synthèse d'image 3D :

Créer une image de synthèse repose sur deux étapes fondamentales : définir la forme tridimensionnelle des objets – c'est-à-dire en donner une description mathématique – et « rendre » l'image, c'est-à-dire calculer une vue du monde virtuel ainsi constitué, prise par une caméra fictive dont on a indiqué la position.²

Il y a donc plusieurs éléments nécessaires à partir desquels le rendu peut se faire : des objets 3D, des matériaux associés à ces objets, une caméra et des lumières . Le fichier numérique d'une scène 3D stocke ainsi toutes les informations relatives à ces différents éléments. Le premier d'entre eux

¹ L'acronyme CGI (Computer Generated Images) est principalement utilisé comme synonyme de la 3D

² M.-P. Cani, « Mondes virtuels », art cit.

repose, on l'a vu, sur une représentation vectorielle de l'espace tridimensionnel. Les « géométries », terme couramment employé pour désigner les objets 3D qui peuplent la scène virtuelle, sont constitués en première instance de points, les *vertices*, à partir desquelles sont tracés les côtés (*edges*) qui, à leur tour, définissent les surfaces. La position de chaque *vertex* est donc définie par un vecteur dans le repère 3D dont les axes x , y et z représentent respectivement la largeur, la hauteur et la profondeur. La géométries sont ainsi visibles dans l'interface d'un logiciel comme *Blender* sous la forme d'un maillage points-côtés-surfaces. Ces surfaces sont souvent constituées de trois ou quatre côtés, les *n-gones* — les surfaces de plus de quatre côtés — étant évités car plus souvent sujets à des *bugs*. De toute manière, au moment du rendu, le moteur opère une *tessellation* qui transforme toutes les surfaces en triangles car c'est la surface minimale — impossible de définir une surface avec moins de trois sommets — donc nécessairement plate et, du même coup, plus simple à calculer.

Une fois qu'on a modélisé un objet 3D, sa surface est nue, souvent recouverte d'une couleur grise uniforme dans le *viewport* des logiciels. Il reste à encore à lui attribuer, dans certains cas, une texture et, surtout, un matériau. La texture, c'est la projection d'une image sur la géométrie 3D : faire correspondre les points de l'images aux points de la géométrie. Un « aplatissement » du volume de l'objet 3D s'impose alors pour le faire correspondre aux deux dimensions de la texture. C'est le principe de l'*UV mapping* qui consiste à découper la géométrie pour l'afficher, comme une sorte de patron, dans un espace bidimensionnel aux axes u et v . Et la texture peut impliquer des images différentes, superposées, codants pour plusieurs paramètres du matériau attribué à l'objet. Si l'image de la *base color* sert à définir la couleur du matériau, les *maps* de *roughness*, de *specular* ou encore de *normal*, permettent de régler le niveau de non-réfectivité, de réfectivité du matériau ou encore de simuler son relief. D'une manière générale, le matériau attribué à l'objet peut être considéré comme l'ensemble des propriétés optiques définissant la façon dont cet objet va réagir à la lumière. Il existe un grand nombre de paramètres configurables selon les matériaux et, lorsque ceux-ci ne sont pas codés directement par des images de texture, ils peuvent être ajustés par des valeurs mathématiques. Dans le cas du *workflow PBR* (*physically based rendering*), l'ajustement des paramètres du matériaux vise souvent l'imitation d'une matière physique. On va, par exemple, augmenter la *roughness* (la non-réfectivité d'une surface) si l'on veut reproduire l'effet du bois ou bien la diminuer pour imiter le métal plus réfléchissant ; on peut utiliser une *map* de *normal* pour simuler les rainures organiques d'un tronc d'arbre ou les aspérités de la rouille, etc... Tous ceci en vue du *shading* : l'opération, au moment du rendu, de la « détermination de l'effet de la lumière sur

le matériau »¹. Fondamentalement, le *shading* c'est donc le calcul de la couleur précise de chaque *pixel* en fonction des propriétés physiques de la surface de l'objet, de sa couleur et de l'illumination de la scène 3D.

Une fois qu'il dispose de la description vectorielle de la géométrie et des informations de *shading* des surfaces, l'algorithme de rendu peut entrer véritablement dans la phase de production d'une image rasterisée. À ce stade, nous sommes encore dans le repère tridimensionnel de la scène virtuel : il faut alors convertir les axes *x*, *y* et *z* du *world space*, espace objectif de la scène 3D, en un espace bidimensionnel, celui du *screen space*. Pour ce faire, le moteur de rendu opère une projection, le plus souvent en perspective, qui permet un « aplatissement » de l'axe *z*. Mais pour cela, encore faut-il un point de vue unique à partir duquel opérer cette projection. C'est à ce moment qu'entre en jeu la caméra virtuelle : comme une caméra physique, elle possède (ou plus exactement simule) des propriétés optiques et, bien sûr, un cadre qui définit les limites de l'image rendue. Surtout, la caméra virtuelle, si elle bien positionnée dans le *world space*, dispose également de son propre espace : le *camera space* ou *view space* dont elle est l'origine, le point de coordonnées (0,0,0) — pour être précis, l'origine exacte du repère se situe au centre du cadre. C'est le basculement du *world space* en *view space* qui permet la projection du repère 3D.

Finalement, une fois que la caméra est en place et que la projection a eu lieu, le dernier élément sur lequel s'appuie l'algorithme de rendu est la lumière. La *rendering equation*, déjà évoquée, décrit l'intensité de lumière qui pénètre dans le diaphragme de la caméra virtuelle en fonction des sources lumineuses de la scène 3D. La force de cette équation c'est qu'elle prend en compte la lumière directe, celle qui pointe vers l'objectif, mais aussi la lumière indirecte qui atteint la caméra après avoir rebondi sur des surfaces réfléchissantes. Dans le monde physique, la lumière provient toujours d'une multitude de sources ; même s'il y a qu'une seule source originale de lumière, ils y ont de nombreuses sources secondaires avec les réflexions. La prise en compte des rebonds lumineux est donc la clé du photoréalisme dont le *ray tracing*, technique d'approximation de la *rendering equation*, offre les meilleurs résultats parmi les différents algorithmes de rendu, ce qui explique que son usage soit devenu la norme pour les moteurs de rendu *offline* employés par le cinéma comme *Renderman* de *Pixar* ou *Arnold* de *Maya*. Le chercheur Thomas Akenine-Möller et ses collègues en offrent une description très claire dans leur ouvrage *Real-Time Rendering* :

¹ T. Akenine-Möller, E. Haines, N. Hoffman, et al., *Real-time rendering*, op. cit., p. 16.

Le *ray tracing* utilise des rayons pour déterminer le transport de la lumière entre les différents éléments d'une scène. Dans sa forme la plus basique, les rayons sont projetés depuis la caméra à travers la grille de pixels vers la scène. Pour chaque rayon, l'objet le plus proche est recherché. Ce point d'intersection est ensuite vérifié pour déterminer s'il se trouve dans l'ombre en projetant un rayon vers chaque source lumineuse et en recherchant si des objets se trouvent entre les deux. Les objets opaques bloquent la lumière, tandis que les objets transparents l'atténuent. D'autres rayons peuvent être générés à partir d'un point d'intersection. Si une surface est brillante, un rayon est généré dans la direction de la réflexion. Ce rayon capte la couleur du premier objet intersecté, dont le point d'intersection est à son tour testé pour les ombres. Des rayons peuvent également être générés dans la direction de la réfraction pour les objets solides transparents, qui sont à nouveau évalués de manière réursive.¹

Pour créer une image rastérisée, le *ray tracing* part donc de chaque *pixel* de l'écran à partir duquel il trace un rayon lumineux qui s'enfonce en profondeur dans la scène 3D. Dès que le rayon rencontre une surface, l'algorithme de *ray tracing* « shade » le point de contact en calculant sa couleur à partir du matériau de la surface, de sa couleur et de la lumière. Le *ray tracing* permet de tracer deux types de rayons : les *primary rays*, les rayons qui partent des *pixels* jusqu'à ce qu'ils heurtent une surface opaque, et les *secondary rays*, rebonds des *primary rays*. De ce fait, le *ray tracing* est très réaliste mais, comme d'autres algorithmes qui reposent sur le même principe de *ray casting* (lancer de rayons), il demande une quantité importante de calcul : « Pour obtenir des images de qualité cinématographique, il peut être nécessaire de tracer des milliards de chemins. »². Pour économiser de la puissance de calcul, le *ray tracing* fonctionne à l'inverse du monde physique : au lieu de partir d'une source lumineuse pour arriver jusqu'à notre œil ou au diaphragme de la caméra, les rayons lumineux sont tracés en sens opposé, depuis la caméra jusqu'à la source lumineuse. Ceci permet de calculer uniquement les rayons utiles pour l'image ; calculer tous les rayons du monde, dont l'infini majorité sont invisibles pour la caméra, demanderait en effet une puissance de calcul infinie. Quoi qu'il en soit, malgré les ruses de *design*, le *ray tracing* est toujours très long. Utilisé par les moteurs *offline* du cinéma, cet algorithme donne lieu à un rendu image par image, loin du temps réel, rapprochant en cela les images de synthèse 3D de l'animation.

¹ *Ibid.*, p. 443-444., Traduction personnelle : “Ray tracing uses rays to determine the light transport between various scene elements. In its most basic form, rays are shot from the camera through the pixel grid into the scene. For each ray, the closest object is found. This intersection point is then checked for whether it is in shadow by shooting a ray to each light and finding if any objects are in between. Opaque objects block the light; transparent objects attenuate it. Other rays can be spawned from an intersection point. If a surface is shiny, a ray is generated in the reflection direction. This ray picks up the color of the first object intersected, which in turn has its intersection point tested for shadows. Rays can also be generated in the direction of refraction for transparent solid objects, again recursively evaluated.”

² *Ibid.*, p. 445., Traduction personnelle : « For cinematic quality images, billions of paths may need to be traced. »

Pourtant, on sait que le rendu *offline* n'est pas une fatalité, le *realtime* des moteurs de jeu vidéo investissant peu à peu le champ du cinéma et son exigence de photoréalisme. Le défi du rendu en temps réel est de taille :

Les modifications du monde calculées à partir de ces données doivent être perçues comme « immédiates », ce qui ajoute une difficulté majeure : le temps alloué pour calculer chaque nouvel état du monde est limité par la fréquence d'affichage ou de retour sensoriel. Ces contraintes, dites de « temps-réel », sont antinomiques de l'exigence de qualité nécessaire à la synthèse d'un monde crédible, voire « réaliste ». Le savoir-faire d'un créateur de mondes virtuels réside dans la réponse qu'il apporte à cette double exigence.¹

La lenteur du *ray tracing* est le prix à payer pour la grande précision de son approximation de la *rendering equation*. De son côté, le *realtime* repose sur un autre algorithme de rendu, bien plus rapide que le *ray casting*, mais dont tout l'enjeu consiste justement à conserver cette rapidité tout en se rapprochant au plus près de la qualité du *ray tracing*. Cet algorithme, est celui de la *rasterization* qui fonctionne en quelque sorte de manière inverse au *ray tracing*. Après la *tessellation*, qui transforme toutes les surfaces de la scène 3D en triangles, puis la projection de ces triangles dans le *screen space*, la *rasterization* détermine, pour chaque *pixel*, quel triangle projeté recouvre ce *pixel*. Ainsi, au lieu de lancer un rayon par *pixel* dans la scène, la *rasterization* part, au contraire, de la scène 3D pour déterminer, pour chaque *pixel*, la surface qui en est la plus proche : celle qui, au premier plan, cache à la vue de la caméra toutes les autres surfaces en arrière. La *rasterization* est donc un algorithme qui classe l'ordre des surfaces par rapport à l'écran. À partir de cet ordre, il est possible de déterminer quelle surface est visible, et quelle autre est cachée, ce qui permet d'économiser une grande puissance de calcul en excluant d'emblée tous les objets invisibles (si un objet est devant l'autre, inutile de traiter l'ensemble des objets en arrière). Et des avancées récentes, notamment au niveau *hardware* avec l'arrivée depuis 2018 des cartes graphiques RTX, permettent désormais un rendu en temps réel hybride, mêlant *rasterization* et *ray tracing* pour certaines tâches plus complexes comme le calcul de la lumière au travers de surfaces transparentes, pour un plus grand photoréalisme des images de synthèse en temps réel.

Quelles que soient les différences entre l'*offline* et le *realtime* concernant le calcul des images, les deux types de rendu s'appuient en revanche sur un même espace semi-persistant : toujours là en tant que description mathématique encodée dans le fichier binaire de la scène 3D

¹ M.-P. Cani, « Mondes virtuels », art cit.

mais présent sous forme d'image rastérisée uniquement dans le cadre de la caméra. On a vu que pour passer de l'espace 3D à l'espace 2D, le moteur de rendu opère une projection à partir de la caméra virtuelle. Le but de cette opération étant de transformer le *frustum* de la caméra — le volume pyramidal dont le sommet est la caméra et qui englobe tout ce qui est dans le champ, le *view volume* — en un *unit cube* dont les points extrêmes sont $(-1,-1,-1)$ et $(1,1,1)$ dans le repère 3D. Ainsi, la pyramide du *view volume* se change en cube, sa base se réduisant à la taille du cadre : dans la projection orthographique tous les objets conservent des dimensions identiques quelle que soit leur distance par rapport à la caméra ; dans la projection perspective, en revanche, les objets les plus loin rétrécissent pour reproduire le phénomène visuel. Tout ce qui, dans la scène 3D, est à l'intérieur du *unit cube* constitue alors l'ensemble de ce qui sera rendu, tout le reste est ignoré. Les géométries à cheval sur les frontières du *unit cube* sont *clipped*, coupées en deux parties, celle qui sera rendue et celle qui retombera dans les limbes de l'espace mathématique. En ce sens, le rendu 3D ressemble à un réalisme teinté d'idéalisme. Dans un moteur 3D, lorsque je « tourne le dos » à un objet, il cesse immédiatement d'exister sous forme d'image rastérisée, tout comme la Lune sort de l'existence dans la conception idéaliste lorsqu'on cesse de la percevoir. Mais bien sûr, il suffit de faire pivoter la caméra pour redonner aussitôt une existence concrète aux surfaces qui se trouvaient en dehors du *unit cube*. Il y a ainsi toujours une forme de réalisme, le hors-champ ne disparaissant jamais complètement puisque le rendu présuppose, non plus une image latente, mais un espace latent décrit mathématiquement et toujours présent dans le fichier de la scène. Le hors-champ est ici caractérisé par sa virtualité plus que par son absence : virtuel numérique en tant que description binaire, mais, surtout, virtuel au sens propre en tant qu'image potentielle actualisée par le rendu, lorsque le hors-champ devient champ. Le rendu 3D met donc en œuvre un espace semi-persistant — persistant en tant que description mathématique mais *mind-dependent*, ou plutôt *view-dependent*, en tant que surface visible. Nous pouvons alors proposer une définition de la prise de vues virtuelles : l'enregistrement de vues dans un espace virtuel semi-persistant dont la 3D est le meilleur représentant.

Un plateau de tournage virtuel

Avant de pouvoir tourner dans un espace virtuel semi-persistant, il faut d'abord le concevoir. Pour l'animation 3D, cela revient à créer un véritable plateau de tournage virtuel, voire un petit monde dans lequel les infographistes peuvent donner libre-court à leur créativité. La création d'une

scène 3D est à la base des différents médias qui utilisent cette technologie (animation, jeu vidéo, VR, architecture, etc...) et nombre de ses étapes leur sont communes. Ce processus constitue une forme étrange de création, impliquant une grande quantité de savoir-faire et de techniques qui empruntent autant au travail du peintre que du sculpteur, du réalisateur évidemment, mais aussi du scénographe, de l'éclairagiste ou encore du *designer*. Surtout, c'est le lieu d'un curieux mélange entre perception et imaginaire : on construit un monde virtuel dans lequel on perçoit et qui, pourtant, n'existait l'instant d'avant que dans notre imaginaire, offrant une expérience de création nouvelle qui entretient une forme d'analogie avec la création divine :

Qui n'a pas rêvé d'être Dieu ? De créer le monde, « son » monde, de le contrôler et de l'explorer librement ? Les nouvelles technologies de l'image nous propulsent dans cette voie. Parfois munis de dispositifs d'immersion – vision en relief, son spatialisé, dispositifs tactiles et même olfactifs, animés pour donner l'illusion de la vie, les mondes virtuels nous permettent d'exprimer librement notre force créatrice.¹

Mais en quoi consiste, concrètement, cette création d'un monde virtuel 3D ? Admettons qu'un artiste veuille créer une scène de bord de mer : quelques baigneurs qui jouent au loin dans les vagues, une plage de sable, des touristes allongés sur des serviettes et, au premier plan, des dunes avec de la végétation. Pour obtenir cette image, l'artiste 3D doit réaliser un certain nombre de tâches successives qui nécessitent, encore une fois, des savoir-faire différents. Après avoir défini son idée – par exemple avec un *concept art*, une illustration de l'image qu'il compte obtenir – il s'attelle au *layout*, c'est-à-dire à une sorte de brouillon de mise en scène dans l'espace 3D. Au moyen d'un logiciel comme *Blender* qui offre un *viewport 3D*, un espace tridimensionnel dans lequel il peut se déplacer et positionner des objets, des lumières ou des caméras, l'artiste met un place une première version de sa scène avec des formes grossières. Dans notre exemple, on peut imaginer quelques bosses pour les dunes, une surface plane pour la mer et quelques rectangles pour l'emplacement des personnages. C'est également à cette étape que les premières intentions de cadrage et de lumière peuvent être signifiées avec le placement dans l'espace 3D d'une caméra et de lampes virtuelles. Ensuite, vient l'étape cruciale de la modélisation : il s'agit de créer, avec le niveau de détail qu'on souhaite, les objets qui vont peupler la scène. C'est en quelque sorte l'équivalent des métiers de concepteur de décors, d'accessoiriste et de costumier qu'on trouve dans le cinéma en prise de vues

¹ *Ibid.*

réelles, avec ceci en plus que, dans le cas de la 3D, il faut également créer de toutes pièces les acteurs. La modélisation peut s'effectuer de différentes manières avec différents logiciels : pour les objets géométriques une souris peut suffire à agencer les différentes arêtes et sommets pour obtenir la forme désirée (un parasol ou une glacière par exemple), en revanche pour les formes organiques, comme un corps humain, il est plus simple d'utiliser une tablette graphique avec un stylet dans un logiciel de sculpture numérique comme *Zbrush* — on ajoute ou on enlève de la matière avec la pointe du stylet qui se transforme en burin virtuel de l'autre côté de l'écran. Si la modélisation et le *sculpt* constituent les techniques les plus répandues, il existe d'autres méthodes pour obtenir des objets 3D, à l'instar du scan que nous déjà évoqué et qui reconstitue la doublure virtuelle d'un objet physique. En parallèle de la modélisation, d'autres étapes peuvent être nécessaires pour la création d'objets particuliers comme les poils, les vêtements ou encore l'eau. Une fois les géométries modélisées, l'artiste leur applique des textures et des matériaux. Nous avons vu qu'il faut, pour cela, découper chaque objet 3D pour en obtenir un patron qui tient dans un espace 2D. L'aplatissement de la surface de l'objet permet de lui assigner une image en guise de texture et il devient alors possible de peindre l'objet. C'est l'étape du *texturing* pendant laquelle l'artiste peint la surface de l'objet 3D dans un logiciel comme *Substance Painter* au moyen d'une tablette graphique et d'un stylet qui s'est mué, cette fois, en pinceau virtuel. Ainsi, dans le cas d'un humain numérique, l'artiste devra peindre la couleur globale et les différents détails de la peau comme les taches de rousseur ou les grains de beauté. Lorsque qu'on dispose des textures pour tous les objets, il faut les appliquer sur la surface 3D avec un *shader*, le programme informatique qui permet de paramétrer le matériau virtuel attribué à l'objet et qui servira plus tard au *shading*. Le *shader* effectue le *mapping* de la texture 2D sur la représentation 3D de sa surface, mais il se charge également de reproduire la façon dont le matériau souhaité réagit à la lumière avec des effets de transparences, de réflexion, d'absorption... Au-delà de cette étape, une chaîne de tâches s'ajoute si, au lieu d'une image fixe, on souhaite obtenir une image en mouvement. Il faut alors « rigger » les objets qui seront animés, c'est-à-dire qu'il faut leur créer un squelette, le *rig*, dont les différents os auront la capacité d'entraîner la surface de l'objet qu'on leur aura assignée. Dans le cas d'un personnage humain, par exemple, on crée une armature qui schématise le squelette humain avec des os pour les jambes, la colonne vertébrale, les bras, les mains, etc... Ce squelette est placé à l'intérieur de la modélisation du personnage et l'on vient peindre sur sa surface les zones d'influence de chaque os : c'est le *skinning*, l'assignation d'une zone de peau à chaque os. Lorsque le *rig* est prêt, on peut animer, et il s'agit alors de bouger le

squelette selon le mouvement qu'on veut obtenir. La séquence d'animation est constituée de poses placées à différents intervalles sur la *timeline*. Pour chaque pose, les os du squelette sont positionnés avec la souris et se voient assigner une « clé » qui les fige à un moment donné. C'est la succession des poses qui crée le mouvement, le logiciel calculant les interpolations entre les poses clés. On a vu qu'il existe d'autres techniques d'animation comme la motion-capture qui consiste à enregistrer les mouvements d'un acteur revêtu d'une combinaison de capteurs, et à appliquer cette séquence de poses au *rig* d'un personnage 3D. Par ailleurs, certains objets ne se prêtent pas à l'animation par le biais d'un squelette, comme les vêtements, les cheveux ou l'eau. Ces animations particulières sont du ressort de la « simulation » qui vise à reproduire de manière procédurale certains phénomènes physiques. Dans notre exemple d'une scène de bord de mer, il faudrait simuler le mouvement des vagues ainsi que l'effet du vent dans la chevelure des personnages, dans leurs vêtements et sur la végétation avec un logiciel comme *Houdini*. Enfin, lorsque toute la matière de la scène 3D a été modélisée ou scannée, c'est-à-dire lorsqu'il en existe une représentation géométrique dans l'espace 3D ; lorsqu'elle a été texturée, qu'on lui a appliqué des matériaux et qu'elle a été animée, cette matière peut être mise en lumière. Concrètement, il s'agit de faire glisser, avec la souris, des lampes virtuelles dans le *viewport* 3D du logiciel. En déterminant de nombreuses options comme la position de la lampe, son orientation, son type (un soleil ? un spot directionnel ?) ou encore l'intensité de la lumière, il est possible d'obtenir l'éclairage souhaité, simulation très proche de l'éclairage qu'on obtiendrait avec la même configuration dans le monde physique. Enfin, on peut placer la caméra virtuelle et se promener dans la scène à la recherche d'angles intéressants. On peut également animer la caméra en lui assignant des poses clés sur la *timeline*, tout comme les *rigs* des personnages, pour lui faire exécuter un mouvement particulier. Une fois qu'on a procédé aux derniers réglages concernant la taille de la focale, l'ouverture, ou la mise au point, il ne reste plus qu'à cliquer sur le bouton « rendu » comme on appuie sur un déclencheur.

Pour le cinéma d'animation, qui a vu progressivement la 3D remplacer la 2D comme technique majoritaire, le processus créatif que nous venons de décrire a constitué une véritable révolution dans l'expérience de réalisation. Si l'animation classique produite par le Walt Disney des années 1930 était avant tout une affaire de papier et de crayon, l'animation de nos jours est principalement informatique : les animateurs qui bougent les squelettes des personnages avec une souris sont désormais bien plus nombreux que ceux qui les dessinent. L'animation à proprement parler — l'étape où l'animateur insuffle le mouvement — représente un cas emblématique du

bouleversement dans la façon de réaliser un film d'animation, mais il n'est pas une étape qui ne soit pas affectée. La mise en lumière, par exemple, change fondamentalement avec la 3D : on a vu, avec le rendu, que le tracé manuel des ombres et des lumières a laissé place aux calculs informatiques qui déterminent le comportement des rayons lumineux émis depuis une lampe virtuelle ajustée par le *lighting artist*. Si le film d'animation 3D est encore proche du dessin animé traditionnel par sa mise en mouvement souvent image par image ou son mode de rendu *offline* toujours majoritairement utilisé, il est plus près du cinéma en prise de vues réelle en ce qui concerne la mise en scène : on travaille dans une scène 3D qui tient beaucoup du plateau de tournage avec un décor, des acteurs, des lumières et des caméras. Le métier de directeur de la photographie qui, pour des raisons évidentes n'existait pas dans le dessin animé, occupe désormais une place importante dans la réalisation d'un film d'animation 3D : comme sur plateau de tournage, il supervise le choix des lumières et des focales de caméra qui exécuteront le rendu. Au point qu'un film comme *L'odyssée de Pi*, constitué en grande majorité d'images de synthèse 3D, a remporté l'Oscar de la meilleure photographie, ce qui aurait été impensable pour un film d'animation avant l'arrivée de la 3D. C'est que ces nouvelles façons de travailler entraînent également une nouvelle esthétique. La 3D permet de réaliser des films d'animation qui n'auraient pas pu voir le jour avec le dessin animé. On pense aux films de *Pixar* dont l'aspect visuel est facilement identifiable et qui offrent une réalisation très proche du cinéma *live* avec un travail poussé sur la lumière et les mouvements de caméra. Si les travellings et autres effets de caméra sont très vite limités dans l'animation dessinée (il faut prendre en compte, pour le dessin de chaque image, le mouvement de la caméra en plus du mouvement des personnages), ils deviennent illimités dans l'animation 3D qui peut même aller beaucoup plus loin que le cinéma en prise de vues réelles. Il est possible de faire exécuter à la caméra absolument tous les mouvements qu'on souhaite : la faire voler, entrer dans la bouche d'un personnage ou traverser des murs. Le film d'animation *J'ai perdu mon corps*, dont les images finales ont été dessinées pour obtenir un aspect graphique 2D, a toutefois fait un grand usage de la 3D pour la mise en scène, le choix des points de vue et des mouvements de caméra : le film entier a été réalisé dans une version brute en 3D avant d'être redessiné, ce qui a permis de combiner l'esthétique de la 2D avec la réalisation immersive et très réaliste de la 3D.

En parallèle de la révolution dans le monde de l'animation, la 3D souffle un vent de changements dans le monde de la prise de vues réelles. Nous avons déjà évoqué la place grandissante des moteurs *realtime* sur les plateaux de tournage avec des techniques comme la

prévisualisation en direct des effets spéciaux ou l'extension du décors physiques avec un *LED wall*. Ces nouveautés technologiques entraînent la formation d'un nombre croissant de directeurs de la photographie aux moteurs temps réel et à leurs espaces virtuels dans lesquels ils retrouvent une expérience assez similaire au tournage. Des chefs opérateurs reconnus comme Greg Fraiser, récompensé d'un Oscar pour son travail sur *Dune* vante les mérites d'un moteur de jeu comme *Unreal Engine* :

J'aimerais vraiment savoir bien utiliser *Unreal Engine*, car si je suivais actuellement une formation de directeur de la photographie, l'outil le moins cher à disposition serait *Unreal Engine*. Avec *Unreal Engine*, vous disposez de *Metahumans*, ce qui vous permet de créer un visage, de l'éclairer et de commencer à explorer le fonctionnement de l'éclairage par le haut. Vous pouvez commencer à comprendre émotionnellement ce que vous ressentez lorsque vous placez une lumière sur le côté, en haut, diffusée ou plus froide. Vous pouvez le faire rapidement. Ils peuvent ensuite appliquer cela au monde réel, mais ils disposent également d'une base de connaissances très positive lorsqu'ils entrent dans le monde virtuel, car ils savent ce qui est possible et ce qui ne l'est pas.¹

Et, en effet, ce qui marque une différence importante entre les *cinematographers* et les infographistes 3D, c'est l'expérience pratique des premiers. Ceux-ci partent d'une expérience de la lumière et de la caméra acquise lors de tournages en prise de vues réelles. Et, de toute manière, l'usage qu'ils font des moteurs de rendu est principalement tourné vers la prise de vues réelles. Sur un plateau de tournage avec un *LED wall*, le directeur de la photographie se retrouve à gérer en direct la photographie dans le monde physique et le monde virtuel à la fois. Cette approche de la 3D par et pour la prise de vues réelles, marque sa différence avec la pratique de l'animation 3D par le terme de *virtual cinematography* ainsi défini :

La cinématographie virtuelle a été définie par l'extension virtuelle d'une expertise pratique en matière de caméras, d'objectifs et d'éclairage, les moteurs de jeu en temps réel constituant la pierre angulaire de la production virtuelle et la prévisualisation devenant un élément incontournable des productions de *blockbusters* en prise de vues réelles.²

¹ Cité dans HOGG Trevor, « Focusing on virtual cinematography », in *VFXV*, 2023., <https://www.vfxvoice.com/focusing-on-virtual-cinematography/>, Traduction personnelle : « I wish that I was really good at using Unreal Engine because if I was training to be a cinematographer right now, the cheapest tool they can get in their arsenal is Unreal Engine. With Unreal Engine, you have metahumans so you can build yourself a face, light it and start to explore how top lights work. You can begin to figure out emotionally how you feel when putting a light on the side, top, defused or cooler. You can do that quickly. They can then apply that to the real world, but also have a huge positive base of knowledge when getting into the virtual world by knowing what is and isn't possible »

² *Ibid.*, traduction personnelle : « Virtual cinematography has been defined by practical expertise on cameras, lensing and lighting while extending its influence, with real-time game engines being the cornerstone for virtual production and previsualization becoming a staple of live-action blockbuster productions. »

La *virtual cinematography* est donc un terme forgé par les chefs opérateurs pour se différencier des animateurs : tous travaillent avec la synthèse d'images 3D mais les premiers ne font assurément pas de l'animation mais toujours de la prise de vues, virtuelles cette fois. Cependant les frontières sont poreuses et, à la croisée de l'animation 3D et de la *virtual cinematography*, se trouve la *previsualization* ou « previs » qui consiste à « tourner » à l'avance en 3D un film destiné à être réalisé, au moins en partie, en prise de vues réelles. La *previs* a pour but de faciliter le tournage de scènes très complexes. Certains studios se spécialisent même en *previs* comme *The Third Floor* ou les *Androids Associés* en France : leur but n'est pas de créer de belles images qui seront intégrées dans le produit fini, mais de fournir une visualisation brute du film avant même le tournage, souvent avec des outils *realtime*. Il est difficile de parler de « brouillon » pour désigner les séquences produites par les studios de prévisualisation car, si les images n'ont pas à être parfaites, devant être réalisées et rendues rapidement, la mise en scène, elle, et en particulier les mouvements de caméra, doit être très précise. Au point que pour certains tournages très complexes, les studios de *previs* fournissent également une *technical visualization* ou « techvis » décrivant la position des objets sur le plateau et les mouvements de la caméra au centimètre près. Ces informations issues du tournage virtuel peuvent même être transmises aux bras mécaniques contrôlant la caméra physique : on a vu précédemment le cas de la caméra physique qui contrôle la caméra virtuelle ; ici c'est l'inverse, les mouvements de la caméra virtuelle sont transmis à la caméra physique qui les exécute avec une précision robotique. Ainsi, pour les films de *live action* reposant en grande partie sur de la *previs* et de la *techvis*, le véritable tournage a plutôt lieu en prise de vues virtuelles. Le film *Gravity*, par exemple, a été entièrement réalisé en *previs* et *techvis* avant d'être tourné : si un tel tournage physique reste tout de même le lieu de l'expression des acteurs de chair et d'os, il ne laisse pas de place pour l'expressivité du chef opérateur ou du cadreur : le tournage a déjà eu lieu.

Phénoménologie de la perception virtuelle

La synthèse d'images 3D permet de créer de véritables simulations de l'espace physique, si précises parfois qu'elles peuvent, comme dans le cas de la *techvis*, servir à contrôler l'agencement du plateau de tournage physique et les mouvements de la caméra. Mais quelle est l'expérience d'un tel tournage virtuel ? Comment la perception s'exerce-t-elle dans ces simulations ?

D'une manière générale, le concept de « perception artificielle » paraît fragile ; il se heurte à plusieurs difficultés. D'abord, il se situe à la croisée d'un double problème que la phénoménologie husserlienne peine à résoudre : le conflit entre perception et *phantasia*, et l'existence sujette à caution de la perception instrumentée. Dans son livre *Le conflit des perceptions*¹, justement consacré à la perception virtuelle, Elsa Boyer revient sur cette « zone de tension au sein de la phénoménologie husserlienne : un flou autour de la frontière entre présentation et représentation »². Tout commence par l'opposition que marque Husserl entre perception et *phantasia*. La perception c'est « la réalité effective actuellement présente », c'est la présentation d'un réel extérieur bien présent. Nous l'avons vu, c'est sur cette présentation immédiate, directe, que repose la prise de vues. D'un autre côté, la *phantasia*, c'est la perception « non originaire », non réellement présente, à l'exemple de l'imagination, du rêve ou du souvenir. C'est sur elle que s'appuient d'autres modes de figuration comme la peinture. Elsa Boyer met en lumière ce point obscur de la pensée husserlienne et pointe le caractère problématique pour la perception virtuelle du conflit des perceptions. En effet, les objets virtuels numériques sont, comme nous l'avons vu, à la fois constitués d'objets numériques, des agrégats de *bits*, mais également d'états mentaux. Leur perception tient-elle alors plus de la perception originaire, présentation, ou de la *phantasia* ? Et le problème s'épaissit encore lorsqu'on rajoute la dimension artificielle de la perception virtuelle. Boyer rappelle qu'Husserl s'était brièvement intéressé à cette question de la perception instrumentée, notamment en étudiant des objets techniques comme le microscope, sans aller cependant très loin. Le mystère demeure : la perception instrumentée est-elle une présentation ou une représentation ? Ce faisceau de questions est clé pour notre sujet puisque, la prise de vues reposant sur la perception originaire et non sur la *phantasia*, si la perception dans les mondes virtuels n'est pas perception originaire, alors nous nous retrouvons dans l'impossibilité de parler de prise de vues virtuelles.

Ce trouble autour de la perception artificielle ne correspond pas à un trouble autour de la réalité même qu'elle donne à voir. On a déjà évacué cette question, et c'était d'ailleurs l'objectif de la phénoménologie que de dépasser l'opposition de l'illusion et de la vérité en faisant de l'apparaître l'être du monde. En revanche, ce qui est en jeu ici, c'est le mode de perception ou plutôt son « degré d'aura phénoménologique » :

¹ BOYER Elsa, *Le conflit des perceptions*, Paris, Éditions MF, « Collection Invention », 2015, 203 p.

² *Ibid.*, p. 4.

Il faut distinguer le degré d'existence d'une chose – en tant que quantum d'être – de son degré d'aura phénoménologique – en tant que quantum de perception. Une chose peut avoir moins d'aura qu'elle n'a d'être, et réciproquement. C'est là qu'est toute la subtilité de la révolution numérique comme révolution ontophanique, et c'est ce qui, longtemps, nous a trompés en nous jetant dans l'illusion du virtuel et la rêverie de l'irréel.¹

Stéphane Vial, dans son livre *L'être et l'écran* consacré à la perception artificielle via le numérique, reprend la notion d' « aura » de Walter Benjamin pour distinguer des degrés d'intensité phénoménologique entre les perceptions. On retrouve très vite la distinction husserlienne entre perception et représentation : la perception originaire étant celle du face-à-face physique, doué du plus grand degré d'aura phénoménologique ; le numérique étant renvoyé, de son côté, à la perception non-originaire, sans doute pas *phantasia* car non imaginaire, mais du moins dotée d'une aura phénoménologique bien plus faible. La conclusion de l'ouvrage de Vial est sans appel : s'il reconnaît au virtuel une certaine aura phénoménologique, celui-ci ne permet pas pour autant des expériences aussi riches que la perception dans le monde physique. Vial prend en particulier l'exemple d'un échange en face-à-face qu'il oppose à un appel en visioconférence. Dans ce cas précis, force est de lui donner raison. Mais toutes les expériences virtuelles ne sont pas à mettre au même niveau : la réalisation d'un film 2D en *puppets* sur *After Effects*, le *lighting* dans un moteur *offline* ou la *virtual cinematography* dans un moteur *realtime* offrent des expériences bien différentes, avec des auras variées ; le temps réel constituant sans doute le stade le plus avancé phénoménologiquement de l'informatique graphique. Cette aura phénoménologique rabougrie que décrit Vial correspond bien à l'idée répandue que la perception virtuelle est désincarnée à l'inverse de l'expérience commune, qu'elle se fait sans corps. Pour donner toute sa chance à la perception virtuelle, examinons le cas de figure où son aura est la plus forte : dans les moteurs temps réel de jeu vidéo.

Contre une description purement formelle du jeu vidéo qui le traite de l'extérieur comme une structure formelle, Mathieu Triclot défend, dans sa *Philosophie des jeux vidéo*², une approche expérientielle du jeu : des *play studies* plutôt que des *game studies*, pour rendre compte de l'expérience du joueur au cœur du jeu — *ce que ça fait au joueur de jouer*. En effet, le jeu constitue bien un « espace intermédiaire », un « lieu » qui est un « entre-deux », ni tout à fait du côté du côté

¹ S. Vial, *L'être et l'écran. Comment le numérique change la perception*, op. cit.

² M. Triclot, *Philosophie des jeux vidéo*, op. cit.

du sujet, ni tout à fait du côté de l'objet, mais plutôt dans la relation, proche en cela de l'expérience du monde vécu et se prêtant particulièrement bien à une approche phénoménologique. Reprenant les travaux du psychanalyste Donald Winnicott, Triclot insiste sur le « contrôle magique » que permet le jeu en dotant le joueur d'une agentivité importante sur ses objets — bien supérieure à celle qu'il peut expérimenter dans la « perception ordinaire » — dont l'existence n'est pour autant pas indexée à sa seule imagination. Ainsi,

L'espace du jeu s'instaure dans une forme de relation magique entre le sujet et l'extériorité, avec des objets extérieurs qui sont en quelque sorte habités, envahis par la subjectivité. Le réel du jeu n'est plus le réel de la perception ordinaire, celui de la complaisance soumise à une réalité extérieure à laquelle il faut s'ajuster et s'adapter. Ce n'est pas encore le réel irréel de l'hallucination ou du délire, dans lequel la réalité extérieure disparaît pour devenir phénomène subjectif. L'état du jeu appartient à un espace potentiel dans lequel une fraction de la réalité de trouve soumise à une forme de contrôle magique, qui maintient l'objet dans un état suspendu, ni complètement ma création ni évènement simplement extérieur. Le jeu instaure des objets étranges, des « objets subjectifs » pour reprendre l'expression de Winnicott.¹

Et, la force du jeu vidéo c'est que, par sa dimension informatique, il actualise avec une force nouvelle le « contrôle magique » du jeu en général, permettant une agentivité débridée du joueur ou du créateur — cela finit par revenir au même — dans le monde virtuel d'un moteur de jeu vidéo. Cette possibilité d'exercer une « action magique », si elle existe depuis les premiers jeux video, portant une responsabilité importante dans le succès du genre, gagne encore en intensité avec l'arrivée de la 3D. Triclot prend l'exemple du genre très populaire du *FPS*, *First Person Shooter*, qui se développe avec l'arrivée des premières cartes graphiques dans les années 90. Le genre, qui repose principalement sur la capacité à se déplacer dans un espace en trois dimensions pour viser des cibles ennemies, offre une expérience immersive absolument nouvelle qui consacre, plus que jamais, la dimension expérientielle du jeu vidéo :

Il y a d'abord dans le jeu de tir un noyau d'expérience spécifique, lié à la machine informatique, une expérience à nulle autre pareil, et que l'on manquera si l'on se concentre sur le versant visible du dispositif, celui que l'on peut observer de l'extérieur, si l'on oublie ce qui se passe dans l'entre-deux, entre le joueur et l'écran.²

¹ *Ibid.*, p. 31.

² *Ibid.*, p. 26.

Surtout, par l'usage qu'il fait de la vue à la première personne — cette vue est l'essence même du genre, comme son nom l'indique — il offre une très grande similarité avec le tournage virtuel. Il s'agit bien à chaque fois de se déplacer dans un espace 3D et d'orienter la caméra pour viser. D'ailleurs, le terme « shooting » en anglais désigne autant le tire que la prise de vues. Le joueur de *FPS* est d'abord un opérateur de caméra : l'essentiel de son expérience de jeu consiste à se déplacer dans l'espace pour cadrer, obtenir des vues privilégiées sur l'action. Qu'en fin de compte il appuie sur la détente d'un fusil virtuel plutôt que sur le déclencheur d'une caméra, la différence est minime par rapport à la similarité de l'expérience. Et, inversement, l'expérience d'un *virtual cinematographer*, qui tourne dans un moteur de jeu, par exemple pour mettre au point une *previs*, est semblable à celle d'un joueur. En effet, si l'expérience du joueur de jeu vidéo et du spectateur de cinéma est assez différente, le second passif, s'identifiant à des personnages, le premier actif, toujours lui-même, l'opposition de tient plus dans le cas, non pas du spectateur, mais du réalisateur. Celui-ci agit dans l'espace, se déplace et vise : lorsqu'il travaille dans des mondes virtuels, son expérience est semblable à celle du joueur de jeu vidéo ; c'est l'expérience du « plaisir inédit, qui est intimement lié à la machine informatique, à la confrontation à un univers engendré par le calcul. »¹. Et, l'action magique permise par cet univers numérique offre des nouvelles possibilités d'expérience à la prise de vues virtuelles empruntées au jeu vidéo : il est possible de faire voler la caméra, de passer à travers des murs, de se téléporter d'un monde à l'autre ou bien de « jouer avec un univers qui offre la possibilité extraordinaire d'une répétition à l'infini. »². Face à ces nouvelles possibilités d'action, Triclot s'interroge alors sur la nature de la perception à laquelle elles donnent lieu. Est-il possible de percevoir dans l'espace virtuel d'un moteur de jeu comme on perçoit dans le monde extérieur, de retrouver quelque chose de la vision naturelle et du sentiment du corps ordinaire dans ces conditions hautement artificielles »³ ?

La réponse négative pour Triclot qui ne voit pas dans la perception virtuelle une expérience instrumentée continuant le corps organique : « Les jeux vidéo sont des jeux étranges dans la mesure où ils se jouent avec une machine, l'ordinateur sous toutes ses formes, plutôt qu'avec un outil ou un

¹ *Ibid.*

² *Ibid.*, p. 25.

³ *Ibid.*, p. 99.

instrument qui prolonge le corps organique »¹. Partant d'une étude, réalisée par Alexander Galloway, de la vue à première personne, emblématique du régime de perception à l'œuvre dans les jeux vidéo, le philosophe conclut à une forme d'« immersion paradoxale, sans corps propre. »². En effet, dans ses *Essays on Algorithmic Culture*, Galloway montre que, contrairement à ce qu'on pourrait imaginer, la vue à la première personne ne permet pas plus d'identification avec le personnage incarné. Au contraire, elle produit toujours une impression de détachement, de perte de repères, et, le cinéaste qui l'utilise doit le faire avec parcimonie, pour des scènes choisies, sous peine de rompre le lien empathique entre les spectateurs et son personnage. C'est que la vue à la première personne n'a pas pour but, dans le jeu vidéo, de générer une identification du joueur avec un personnage mais, au contraire, de lui laisser une place vide qu'il peut occuper : « En jeu vidéo, il n'y a pas de personnage (principal), il n'y a que des avatars, c'est-à-dire des marionnettes, des places vides pour une logique de l'action plutôt que du regard. »³. La logique du jeu vidéo est donc bien différente de celle du cinéma, non pas générer de l'empathie mais de l'action :

Autrement dit, le même dispositif de vue subjective intègre un autre rapport à l'image, où il ne s'agit plus prioritairement de produire des émotions et de véhiculer une histoire à travers des tâches de lumière sur un écran, mais d'expérimenter un univers actionnable et qui appelle l'action. La vue subjective nous enjoint en quelque sorte à nous déplacer. Elle nous incite à recréer l'espace habitable, à quitter le point fixe, à s'arracher à la vision mutilée du simple spectateur. Le jeu vidéo commence au point exact où le cinéma s'arrête, avec une forme d'image-action qui appelle le mouvement plutôt que le regard.⁴

Cette « désincarnation » que décrit Triclot est sans doute valable pour le spectateur, mais l'est-elle pour le joueur ou l'opérateur caméra ? Au contraire, l'« image-action » constituée par la vue à la première personne, loin d'instituer une immersion « sans corps propre » semble plutôt en appeler, plus qu'aucune image avant elle, au corps. La conclusion de Triclot semble en effet contradictoire au regard de ce que l'on sait sur l'expérience de l'espace telle que la phénoménologie, en particulier celle de Merleau-Ponty, la décrit. On a vu que, chez le phénoménologue, le corps est la condition de l'expérience spatiale, le point de « pivot » à partir duquel s'exerce la dynamique de la perception et

¹ *Ibid.*, p. 39.

² *Ibid.*, p. 100.

³ M. Triclot, *Philosophie des jeux vidéo*, op. cit.

⁴ *Ibid.*, p. 102.

de l'action. Or, la description que livre Mathieu Triclot de l'image-action du jeu vidéo semble parfaitement correspondre à cette dynamique ; d'ailleurs il parle d'« espace habitable » faisant écho à l'habitation du monde décrite par Merleau-Ponty. Ici, Triclot semble tomber dans un raisonnement trompeur qui voudrait que de l'immatérialité des simulations virtuelle on conclut à une perception sans corps. Or, encore une fois, le corps est la condition de l'expérience, de toute expérience, et la perception virtuelle n'y échappe pas. D'abord, le joueur de *fps* ou le *virtual cinematographer* visent et choisissent leurs cadrages avec leur vision. Mais, si c'est le plus évident, c'est loin d'être le seul sens coporel en jeu. La proprioception est également fortement mobilisée : la 3D engage la capacité à se déplacer, à éviter les obstacles et à tourner autour des objets pour découvrir leurs faces cachées. C'est une véritable prolongation de la mobilité du corps : lorsque je contrôle la caméra dans un environnement virtuel, je suis bien dans une spatialité de situation et non de position. Le point de vue à la première personne, que Triclot décrivait à juste titre comme une « place vide » que le joueur peut occuper, est justement la condition de cette extension de la perception incorporée, de cette perception instrumentée. L'avatar ou la caméra que je contrôle dans un moteur de jeu sont toujours moi, ce sont des ambassadeurs de ma subjectivité dans l'espace virtuel. Contrairement à ce qu'il croit, Triclot fait en vérité, avec la vue à la première personne, la démonstration de l'incarnation de l'expérience perceptive des mondes virtuels. Il explique que la vue *pov* (*point of view*) focalise la perception du joueur sur l'écran et sur le jeu. Or, si la perception du monde extérieur est atténuée, ce n'est pas le signe d'une perte de corporalité, le joueur s'abîmant dans l'immatérialité du virtuel, mais plutôt l'effet de la prolongation de l'appareil perceptif du joueur dans le monde virtuel. La caméra virtuelle constitue, pour celui qui la manipule, un appendice de son corps propre, comme l'est bâton pour l'aveugle. C'est justement sur l'action à laquelle incite la vue à la première personne et qui est permise par l'interactivité des moteurs de jeu, que repose l'acquisition de l'espace virtuel pour le joueur comme un véritable espace, espace d'abord pour le corps propre avant de l'être pour le corps objectif. On se souvient en effet que, chez Merleau-Ponty, le corps est d'abord dans une spatialité de situation qui le place comme origine absolue à partir de laquelle tous les objets de l'action trouvent une position. Ainsi, dans cet état premier,

mon corps m'apparaît comme une posture en vue d'une certaine tâche actuelle ou possible. [...] Le mot « ici » appliqué à mon corps ne désigne pas une position déterminée par rapport à d'autres

positions ou par rapport à des coordonnées extérieures, mais l'installation des premières coordonnées, l'ancrage du corps actif dans un objet, la situation du corps en face de ses tâches.¹

Et c'est seulement par l'action, au moment même où elle se produit, que le corps peut se retourner sur lui-même, « se ramasser » pour constituer le corps objectif, cette fois dans une spatialité de position. Or ce phénomène de ramasse ment du corps, cette dynamique du corps propre et du corps objectif, on la retrouve au moment du rendu avec le basculement du *world space* au *view space*. La caméra qui, jusqu'à là occupait une position donnée dans le repère mathématique de la scène, position à laquelle l'avait placé l'opérateur, se retrouve brusquement, au moment de la projection, dans une spatialité de situation. Soudain, elle est l'origine absolue du repère, le point (0,0,0) à partir duquel tous les objets de la scène se voit attribuer une position. Le *clipping* et la mise de côté du hors-champ au moment du rendu font ainsi écho à l'idéalisme du corps phénoménal qui ne voit toujours qu'un seul côté des choses à la fois et à qui il est permis de douter que la face cachée de la Lune existe.

Il est alors possible pour l'utilisateur des mondes virtuels générés par les moteurs de jeu, de les habiter, de s'installer en eux en exerçant, avec ses sens corporels, « cette relation vitale avec le monde qui nous le rend présent comme lieu familier de notre vie »². Bien sûr, certains sens offrent encore une expérience limitée. Si la vue, l'ouïe ou la proprioception sont bien représentés, c'est plus difficile pour l'odorat et le goût qui sont les grands absents. Le toucher, quant à lui, est ambivalent : on peut le provoquer par synesthésie, mais son manque se fait toujours sentir, introduisant parfois un décalage désagréable dans certaines expériences immersives, par exemple lorsqu'un film en *VR* nous donne l'impression visuelle de marcher sur un parterre organique mais que nos pieds ne rencontrent que la dureté lisse d'un sol de béton. Quoi qu'il en soit, les moteurs 3D constituent bien un type d'expérience instrumentée prolongeant le sentir du corps. Ici, la prise de vues virtuelles ressemble à la prise de vues réelles. Comme l'opérateur de caméra qui, l'œil collé au viseur, finit par l'oublier, l'opérateur virtuel oublie très rapidement la souris et tout autre dispositif d'interaction pour s'immerger complètement dans l'espace virtuel. À ce moment là, l'expérience virtuelle peut même être conçue comme une augmentation de l'expérience du monde vécue : j'ai accès à d'autres modes de déplacements, à des super pouvoir grâce au « contrôle magique » de la matière calculée... Finalement, la perception virtuelle apparaît comme la meilleure réalisation à ce jour du rêve des

¹ M. Merleau-Ponty, *Phénoménologie de la perception*, op. cit., p. 115-116.

² M. Merleau-Ponty, *Phénoménologie de la perception*, op. cit.

cybernéticiens d'aboutir à une symbiose homme-machine. Mais, pour aller jusqu'au bout, cette symbiose doit permettre le développement d'un véritable être-au-monde.

Être-au-monde virtuel

L'être-au-monde, c'est l'expérience du monde vécu dans un corps, c'est l'expérience première à partir de laquelle peuvent se déployer toutes les autres. Chez Merleau-Ponty, cette notion occupe une place centrale, en particulier dans la *Phénoménologie de la perception*. L'être-au-monde concerne non plus l'essence mais l'existence qui tient lieu de conscience incarnée, ou plutôt incorporée, au sujet de la perception : « [...] nous substituons, comme sujet de la perception, à la conscience l'existence, c'est-à-dire l'être au monde à travers un corps »¹. Être-au-monde c'est donc d'abord exister comme sujet de perception, et, nécessairement, comme corps. Cette existence implique une sortie permanente de soi pour aller à la rencontre du monde, le toucher comme dans l'ontologie du Sensible de Barbaras. C'est l'occasion d'une découverte permanente, d'une confrontation avec les choses qui peut donner lieu à des expériences douées d'une forte aura phénoménologique :

Car exister ou être-au-monde, c'est également savoir apprécier les instants non humains, les moments de rencontre avec les faits, les occasions de promenade parmi les choses. Vivre, c'est aussi vivre avec les choses et savoir goûter leur aura phénoménologique.²

Nous n'avons d'ailleurs pas le choix de cette rencontre qui nous est imposée du fait même de notre condition corporelle : l'être-au-monde implique toujours un être-jeté — nous sommes jetés dans le monde au moment de notre naissance. Le monde apparaît alors comme ce déjà-là englobant qui constitue le fond ultime de notre expérience, l'horizon toujours en vue et toujours indépassable. Pascal Dupond, dans son *Dictionnaire Merleau-Ponty*³, distingue plusieurs particularités du concept de « monde » chez le phénoménologue. D'abord, le monde est une unité inter-subjective : il n'est pas multiplié par la pluralité des points de vue ou des expériences ; au contraire, c'est un fond commun à toutes ces expériences. Ensuite, le monde est une totalité : ce qui fait qu'une chose, une

¹ *Ibid.*, p. 357.

² S. Vial, *L'être et l'écran. Comment le numérique change la perception*, op. cit.

³ DUPOND Pascal, *Dictionnaire Merleau-Ponty*, Paris, Ellipses, « Dictionnaire », 2007.

œuvre, une époque, peuvent être « un monde » c'est qu'elles comportent une multitude de dimensions qui peuvent se rapporter, par analogie, aux dimensions du Monde. Ainsi, Merleau-Ponty relève-t-il dans les notes de son ouvrage inachevé *Le visible et l'invisible*, qu'un monde au sens figuré – « c'est tout un monde, le monde du son, de la couleur, etc... »¹ – est « un ensemble organisé qui est clos mais qui, étrangement, est représentatif de tout le reste, possède ses symboles, ses équivalents pour tout ce qui n'est pas lui[...] »². Puis, le monde, c'est bien sûr, le fond unique de notre expérience : il n'y a pas comme chez Leibniz une infinité de mondes possibles, ou, comme chez Descartes, un seul monde réel contre les autres :

Avec Husserl, l'unicité du monde signifie non qu'il est actuel et que tout autre monde est imaginaire, non qu'il est en soi et tout autre monde pour nous seulement, mais qu'il est à la racine de toute pensée des possibles, qu'il s'entoure même d'un halo de possibilités qui sont ses attributs [...]³

Enfin, le monde est un « ouvrage inachevé » : contre l'objet scientifique qui est une totalité finie, le monde est au contraire un ensemble toujours en train de se faire, « une multiplicité ouverte et indéfinie »⁴. Cette dernière caractéristique rejoint la théorie des esquisses et le Sensible de Renaud Barbaras : le monde est toujours fuyant, ne se laisse jamais saisir complètement. Finalement, c'est même à cet apparaître permanent qu'on le reconnaît :

La synthèse qui me donne l'objet sur fond de monde est ouverte, inachevée, sans que cet achèvement prive la chose ou le monde de réalité ; [...] le réel se retire dans sa donation et se donne dans son retrait, et le monde est l'un des noms de ce lien insécable de la donation et du retrait.⁵

Pour habiter le monde, il faut donc le vivre en corps, exercer ses sens en son sein. Nous avons vu, dans le chapitre précédent, que la perception virtuelle dans un moteur de jeu était bien une véritable perception, perception originaire incorporée. Mais, si elle permet l'acquisition d'un espace, cette

¹ MERLEAU-PONTY Maurice, *Le visible et l'invisible: suivi de Notes de travail*, Paris, Gallimard, « Tel », 2016, 359 p, p. 277.

² *Ibid.*

³ *Ibid.*, p. 281-282.

⁴ M. Merleau-Ponty, *Phénoménologie de la perception*, *op. cit.*, p. 85.

⁵ P. Dupond, *Dictionnaire Merleau-Ponty*, *op. cit.*, p. 149.

perception virtuelle permet-elle, du même coup, celle d'un monde ? En d'autres termes, un être-au-monde virtuel est-il possible ?

Tout d'abord, la perception virtuelle permet bien d'habiter l'espace virtuel dans le sens où le couple de l'activité perceptive et de l'activité motrice peut s'y déployer comme dans l'espace physique. Nous percevons donc dans cet espace grâce à notre corps ; la perception virtuelle est une perception instrumentée qui prolonge notre subjectivité. Pour autant, on ne peut pas dire que nous soyons *au monde* virtuel comme nous sommes au monde physique. Dans le cas de l'animation 3D, nous avons vu tout le processus de construction nécessaire de la part du réalisateur ou de son équipe : le monde virtuel ne nous précède pas car c'est nous qui le créons de toutes pièces. Nous ne sommes donc pas, ici, *jetés dans le monde* virtuel comme nous le sommes dans le monde extérieur. Mais, il ne s'agit pas, de toutes manières, de remplacer le Monde, fond unique de toutes les expériences possibles, par un monde virtuel, ce qui serait une contradiction — les mondes virtuels des moteurs de jeu faisant justement partie du Monde. Il est plutôt question de savoir si l'on peut envisager une simulation 3D comme un monde plutôt que comme un simple espace, c'est-à-dire de déterminer si elle constitue un système assez complet pour que ses différentes dimensions permettent de soutenir une analogie avec le concept général de monde. Examinons alors les mondes virtuels des moteurs de jeu sous le prisme des critères mondains que nous venons d'évoquer. Le premier — le monde comme unité intersubjective — ne s'applique pas à la scène 3D d'un film d'animation. En effet, il n'y a, dans ce cas, qu'un seul point de vue, celui de l'opérateur unique, seul dans la scène (d'ailleurs une scène ne peut toujours être ouverte que par une seule personne à la fois sous peine de causer des conflits de fichier). Une telle scène est ainsi, sous ce premier critère, un simple espace qui n'est habité par aucune autre subjectivité que celle l'utilisateur unique. On peut, en revanche, évoquer d'autres exemples fournis par les jeux vidéo, où des joueurs incarnant des avatars dans un univers virtuel donnent bien lieu à un monde intersubjectif. Dans un tel cas, le premier critère est rempli. Le second — le monde comme totalité — fonctionne déjà mieux avec le cas d'une scène 3D de film d'animation. On a vu à quel point une telle scène fonctionne déjà comme un petit plateau de tournage complet avec objets, décors, personnages, lumières et caméra. En ce sens, il y a déjà une forme de totalité. Par rapport à la peinture qui n'est qu'une surface, une scène 3D est une reproduction en profondeur. D'ailleurs, le propre de la synthèse d'image c'est d'être une création par intermédiaire de la machine : au lieu d'être la création de l'image elle-même, c'est la création d'un système, une description mathématique, qui permettra de rendre l'image de

manière autonome. On peut également, encore une fois, évoquer l'exemple du jeu vidéo comme niveau supérieur d'autonomie par rapport à la scène d'animation 3D. Certains jeux comme les *Sims* ou *GTA* fonctionnent justement comme des totalités, des mondes complets. Le troisième critère — le monde comme fond unique de notre expérience — nous renvoie vers la perception virtuelle. Nous avons vu qu'en tant que perception instrumenté prolongement notre appareil perceptif elle plonge, en quelque sorte, notre corps dans le monde virtuel. Bien sûr, nous continuons de percevoir dans le monde extérieur, mais, la perception virtuelle, qui s'exerce avec la vue à la première personne, permet justement une focalisation importante de la perception à l'intérieur du monde numérique. Le joueur déjà concentré sur l'écran, voit son attention encore plus canalisée par la vue subjective. La logique de l'absorption de l'attention par l'écran est alors renforcée par la 3D des moteurs de rendu, permettant un haut niveau d'immersion : à ce moment précis d'immersion, le monde virtuel constitue bien l'horizon principal, si ce n'est unique, de notre expérience. Enfin, le quatrième critère — le monde comme ouvrage inachevé — est au cœur des enjeux de réalisme des simulations 3D, les éditeurs de jeu rivalisant d'ingéniosité pour concevoir des « mondes ouverts » donnant aux joueurs la sensation d'une possibilité quasiment infinie d'exploration. Ce critère d'inachèvement est sans doute le plus exigeant, et il n'est pas rempli, au moins en apparence, par la grande majorité des jeux. Nous avons vu la complexité du processus pour créer une scène 3D complète avec des personnages, un décor et des interactions possibles. On comprend alors que tout soit déjà conçu d'avance, préexistant : en ouvrant les possibilités d'interaction avec un monde virtuel, on augmente la charge de travail de manière exponentielle. Les jeux vidéo qui se vendent comme des « mondes ouverts », à l'instar de *No man's sky* qui offre la possibilité d'explorer un nombre infini de planètes, s'appuient sur des architectures procédurales qui produisent des variations à l'infini à partir des mêmes éléments. Il s'agit donc de recombinaisons plutôt que d'une véritable génération, même si l'effet est convaincant. Voilà pour les difficultés qu'il y a à concevoir des mondes virtuels véritablement ouverts, véritablement infinis. En revanche, au niveau de la perception virtuelle on retrouve bien le principe de la donation par esquisse sur laquelle s'appuie l'inachèvement du monde phénoménal. À l'inverse du film déjà rendu, fichier stockant toujours la même séquence d'images, le propre du programme en temps réel, de la scène 3D ou du jeu vidéo, c'est d'offrir une expérience calculée sur le moment, en fonction des actions de l'utilisateur ; donc une expérience toujours nouvelle, jamais épuisée. Georges Lucas résume ainsi l'apparaître

permanent d'un monde virtuel : « Le jeu vidéo, c'est du cinéma. Seulement, on finit par tourner dix films différents. On a toutes ces options et, à chaque fois qu'on joue, on joue avec le spectacle ». ¹

Au regard de ces différents critères, il semble que les jeux vidéo soient bien plus proche d'offrir l'expérience d'un monde, d'être le lieu d'un être-au-monde, que les scènes 3D dans lesquelles travaillent les infographistes réalisant des films d'animation. C'est que les jeux vidéo sont bien plus proches du modèle de la simulation dans les capacités interactives qu'ils comportent. À l'inverse des scènes 3D pour l'animation qui ne sont accessibles qu'aux créateurs, les spectateurs n'ayant accès qu'au film rendu, les mondes virtuels des jeux vidéo doivent être actionnables par l'ensemble des joueurs, et offrir l'expérience interactive la plus complète, donner l'apparence de l'autonomie, voire de la vie. Conscient de la richesse des expériences que peut procurer le jeu vidéo, des réalisateurs-joueurs s'en sont emparés comme plateaux de tournage. Mathieu Triclot retrace le fantasme du cinéma de représenter le jeu vidéo, fantasme qui remonte au moins au film *Tron*, sorti en 1982. Mais à cette époque, le cinéma joui encore d'une supériorité visuelle par rapport aux limitations techniques du jeu vidéo. Triclot décrit le changement introduit par l'arrivée du *CD-ROM* qui augmente la capacité de stockage des jeux et leur permet, désormais, d'inclure des séquences de film en leur sein, les cinématiques, qui brouillent un peu plus les frontières entre cinéma et jeu :

D'un côté, avec la cinématique, nous avons affaire à des bribes authentiques de cinéma dans le jeu, qui bénéficient d'ailleurs souvent des savoir-faire d'Hollywood ; de l'autre, avec la 3D, c'est le jeu lui-même qui se mue dans son ensemble en un genre entier de quasi-cinéma.²

La 3D transforme le jeu vidéo en un genre de cinéma « dont vous êtes le héros »³ et, de fait, les joueurs deviennent du même coup des réalisateurs. Nous avons déjà évoqué le parallèle entre les jeux de tir *FPS* et la prise de vues, les deux cas de figure impliquant un contrôle subjectif de la caméra. Poussant l'analogie encore plus loin, certains joueurs se mettent à enregistrer la partie de jeu, et à monter des films à partir de ces enregistrements. Ces films « tournés » dans des jeux vidéo ont un nom : les *machinimas*. Si certains réalisateurs de films *live* comme Spielberg ou Gus Van Sant se sont déjà servis de jeux pour prévisualiser leur tournage, ici, l'image finale est rendue par le

¹ Cité dans M. Triclot, *Philosophie des jeux vidéo, op. cit.*, p. 81.

² *Ibid.*, p. 88.

³ *Ibid.*, p. 93.

moteur de jeu lui-même et capturée en temps réel sur l'écran. Le genre se popularise au point que certains jeux comme *GTA* intègrent même un « mode réalisateur » permettant un meilleur contrôle de la caméra et des lumières virtuelles. Le film *Martin Pleure* de Jonathan Vinel, réalisé dans *GTA*, constitue un bel exemple du genre, tout comme le documentaire *Knit's island* déjà mentionné. Dans le cas d'un tel tournage, les réalisateurs sont véritablement *jetés au monde* virtuel : ils *spawnent* nus dans la *map* et doivent survivre, apprendre à se nourrir, à trouver refuge ou à se défendre s'ils veulent « rester en vie » et continuer à tourner. Il faut alors engager une vraie relation avec le monde, faite de compromis et d'actions décidées. Le fait que *Knit's Island* soit un documentaire est particulièrement révélateur : on documente ce qui existe déjà, hors de nous, ce qui nous précède, contrairement à la fiction qui implique une part plus ou moins grande, mais toujours là, de construction. En ce sens, le documentaire exploite au maximum les possibilités de la prise de vues dans ce qu'elle a de spontané et de surprenant. La possibilité de documenter un tel espace virtuel constitue ainsi la meilleure preuve qu'il s'agit bien d'un véritable petite monde.

Pour autant, même dans un jeu vidéo très poussé et complet comme *DayZ* qui vise, justement, l'illusion d'un monde ouvert cherchant à faire oublier sa dimension simulée, le joueur-réalisateur se heurte constamment à des limitations qui peuvent altérer l'immersion. Triclot décrit ainsi la période de cartographie des limites du jeu qui a lieu à chaque début de partie :

Il se produit toujours, au moins au début des jeux, un phénomène d'agencement du joueur au jeu, qui vise à fabriquer le terrain, l'espace dans lequel les interactions positives et gratifiantes pourront avoir lieu. Plus un jeu est riche, plus il propose un monde complet et vaste, plus le phénomène d'ajustement prend de l'ampleur. [...] Pendant quelques heures, au début de la partie, le joueur pourra encore hésiter sur les possibles du jeu : est-ce que telle porte s'ouvre ou bien est-ce que les bâtiments ne se réduisent qu'à leur façade ? Est-ce que je peux parler à tel ou tel personnage ou bien n'ai-je affaire qu'à des pantins qui me renverront toujours la même ligne de dialogue stéréotypée ? Il y a ainsi tout un moment initial du jeu qui est dédié à la découverte et à la cartographie des limites du terrain, celui de l'interaction satisfaisante.¹

Le réalisateur de *machina* doit effectuer ce travail de repérage en prenant garde à ne pas tomber dans les crevasses, nombreuses, de la simulation. Et ces limites auxquelles il se heurte, il va sans dire que le réalisateur d'animation 3D les rencontre multipliées au centuple. Puisque dans les films seul compte le champ, on ne fabrique généralement que ce que le *storyboard* a prévu de montrer, on

¹ *Ibid.*, p. 34.

n'anime que les mouvements nécessaires au scénario, on ne simule que les phénomènes physiques qui termineront dans le montage final. Il est donc d'autant plus difficile pour le réalisateur d'animation de faire de la prise de vues virtuelles. Là où le réalisateur de *machinima* prend le risque, en cherchant à ouvrir une porte, qu'elle ne soit qu'un décors de façade, du moins lui est-il permis d'espérer. Le réalisateur d'animation 3D n'ose, parfois, même pas déplacer la caméra parce qu'il sait que tout ce qui existe du monde virtuel est déjà dans le champ. La possibilité de la prise de vues virtuelles est donc conditionnée à l'existence d'un monde virtuel assez riche pour permettre la sérendipité, les heureux accidents du tournage. Les réalisateurs de *machinimas* possèdent un avantage certain, en ce domaine, par rapport à ceux de l'animation puisqu'ils n'ont pas à concevoir le monde dans lequel ils vont tourner. Ils ont d'emblée accès, en tant que joueurs, à des univers qui peuvent être d'une très grande richesse ; permettant de longs tournages et l'accumulation de *rushes* en vue d'un montage, chose rare pour l'animation dont, on l'a vu, le *storyboard* tient lieu de montage. De toute manière, les mondes virtuels des moteurs de rendu n'auront jamais la complexité du monde extérieur, et la prise de vues virtuelles devra compenser cette simplicité relative par une plus grande ouverture aux découvertes et aux accidents, par la recherche d'un *être-au-monde virtuel*. Les réalisateurs de *machinimas* l'ont bien compris qui ne se laisse pas effrayer par les défauts esthétiques, les problèmes de chargement de textures et autres *bugs*. Les incongruités du comportement de certains avatars, transformés en acteurs principaux, fait aussi la richesse du genre qui n'a pas finalement de raison de faire oublier l'origine de ses images. De la même manière, l'animation 3D pourrait prendre ses distances avec le photoréalisme et le figurativisme, accueillir les *bugs* et les « trous » dans le décor virtuel : elle y gagnerait certainement en flexibilité et en plaisir de création. C'est la promesse de la prise de vues virtuelles.

VI. Expérience de la génération d'images

Générer des films

L'idée d'un art généré entièrement par les machines n'est pas nouveau. Déjà, dans les années 60 l'artiste Harold Cohen avait mis au moins le système *AARON*, un petit robot couplé à un programme lui permettant de « dessiner ». D'abord seulement capable de tracer des formes très simples en noir et blanc que Harold Cohen complétait en peignant à la main, ses capacités se sont enrichies à mesure des améliorations apportées par l'artiste qui continua à travailler sur son programme tout au long de sa vie. Au cours des années 80, *AARON* se tourne vers le dessin figuratif, réalisant des motifs organiques puis des personnes dont les silhouettes élancées ne sont pas sans rappeler celles d'Egon Schiele. Finalement, au tournant des années 2000, le programme acquière la capacité de peindre en couleur, progressant vers une forme d'autonomie jusqu'à la mort d'Harold Cohen en 2016. Celui-ci aura construit une œuvre artistique d'une très grande originalité dans laquelle il est difficile de séparer création technique et création artistique, outil et résultats. Cohen ne se voyait pas comme un programmeur mais comme un artiste travaillant en collaboration avec sa machine. *AARON* constitue sans doute l'un des premiers cas de génération d'image. Mais il ne s'agit pas d'une IA générative : le programme n'est pas en mesure d'apprendre seul de nouvelles variantes et chaque fonctionnalité doit être explicitement codée. Ce fonctionnement implique donc de découper un dessin en un ensemble de formes symboliques, permettant ensuite de coder l'acte de dessiner comme une suite d'actions précises.

La véritable générativité vient avec le changement de paradigme du *deep learning* : il ne s'agit plus, comme dans le cas de l'IA symbolique, d'écrire un algorithme décrivant une suite d'actions, mais d'entraîner un programme pour qu'il soit capable de réaliser des actions non-explicitement codées. En 2015, *Google* lance le réseau *DeepDream* qui fait parler de lui en générant des images psychédéliques semblables à des hallucinations. *DeepDream* devient très populaire et fait connaître au grand public la génération d'images. Plusieurs architectures de modèles se développent et de nombreux artistes s'emparent de ce nouveau médium. En 2017, l'artiste turque Memo Atken conçoit l'installation *Learning to see*, qui traite en temps réel des objets capturés par une caméra et leur applique différentes texture comme des motifs floraux, de feu... En 2021, *Open AI* lance *DALL-E* dont les résultats contrastent fortement avec ceux de *DeepDream* par le réalisme

figuratif de ses images. Si le succès de *DeepDream* tenait au caractère halluciné et incontrôlable des images produites, *DALL.E* semble ouvrir la voie à l'utilisation effective de l'IA générative par l'industrie créative en offrant des possibilités de contrôle au premier rang desquelles le *prompt*, une description textuelle guidant la génération. L'année suivante, *Midjourney* marque une nouvelle rupture et fait les gros titres avec un épisode singulier : l'artiste américain Jason M. Allen remporte un concours de peinture avec une image générée par le modèle sans que le jury ne soit au courant de la technique utilisée. Cet événement entraîne un débat passionné autour de la place de la génération d'images dans l'art, de création véritable et inspiré à plagiat sans âme. En parallèle, apparaissent les premiers systèmes de génération de vidéo. Les débuts sont balbutiants, les modèles peinant à maintenir une consistance temporelle tout au long des séquences générées : ils se contentent, dans les premiers temps, d'assembler une suite d'images produites avec les modèles classiques. Ceci a pour effet de donner des vidéos « clignotantes », instables, suite ininterrompue de métamorphoses qui ne se fixe jamais sur un sujet précis. Les vidéos générées avec *Stable Vidéo Diffusion* autour de 2022-2023 constituent de bons exemples de ce manque de persistance temporelle : dans une vidéo de dix secondes, les figures et les décors évoluent en permanence, s'interpolent offrant des visions surréalistes... Mais cette instabilité ne dure qu'un temps, et un nouveau cap dans le réalisme est franchi en 2024 avec l'arrivée de nouveaux modèles comme *Sora* d'*Open AI*, *Runway Gen-3*, ou encore *Pika Labs*, qui apportent des améliorations considérables dans la persistance temporelle. À partir d'une simple *prompt*, il est possible de générer une séquence d'une dizaine de secondes avec un niveau de photoréalisme qui rend parfois difficile la distinction de la prise de vues réelles. À la date d'écriture, la génération de vidéo a atteint encore un nouveau pallier avec *Runway Gen-4* et, surtout, la très récente sortie de *Google Veo 3* capable de générer des vidéos d'un réalisme exceptionnel, notamment dans la gestion de phénomènes physiques comme le comportement des fluides, ainsi que d'une bande son les accompagnant. Le cinéma semble plus que jamais à la portée de l'IA générative.

Et, de fait, le monde du septième art s'intéresse de très près aux révolutions techniques successives de l'intelligence artificielle dont l'influence étend son ombre sur les plateaux de tournage. La grève de 2023 des syndicats d'acteurs et de scénaristes américains, motivée en grande partie par l'essor de l'usage de l'IA générative par les studios, constitue l'une des manifestations des bouleversements à l'œuvre. L'annonce, par plusieurs cadres de studios hollywoodiens, d'un usage de l'IA pour baisser les coûts ou faciliter les productions attise les tensions. Certains acteurs militent

ainsi pour ne pas être remplacé par des doublures numériques quand d'autres demandent simplement l'actualisation d'un droit à l'image portant sur les *deepfakes*. Et le combat des scénaristes est semblable : à défaut de rendre totalement hermétique le champ de l'écriture à ces nouvelles technologies, du moins garantir une rétribution des auteurs dans le cas où leurs textes serviraient de base d'entraînement ou de travail à des IA. L'enjeu est de taille car, de la génération d'images à la génération de vidéo en passant par le texte ou le son, aucune des étapes de la production d'un film ne semble hors d'atteinte de l'IA générative. Concernant la préproduction, les outils se multiplient : des modèles comme *Cynelitic* fournissent des prédictions de rentabilité d'un film à partir d'un simple scénario ; des outils comme *Sudowrite* permettent de construire des arcs narratifs, d'étoffer des personnages et même de générer des scénarios complets ; sans compter, bien sûr, les modèles de génération d'image comme *Midjourney*, de plus en plus intégrés à la production des *concepts arts*, des *storyboards* et *animatics*. La startup *LTX studio* propose même une suite complète de modèles couvrant toutes les étapes, depuis l'écriture jusqu'à la génération de rushes en passant par le storyboard. Au cœur de la production, la création de séquences vidéo constitue sans nul doute le gros du morceau, et les studios suivent de près les avancées des modèles de *text-to-video*, certains allant même jusqu'à développer des partenariats avec des entreprises technologiques, à l'instar de celui qui lie *Lionsgate* et *Runway*. Avec ce partenariat, *Lionsgate* donne accès à l'ensemble de son catalogue à *Runway* pour entraîner un nouveau modèle sur mesure à l'usage exclusif du studio. En les producteurs ne sont pas les seuls à frayer avec le secteur de la *tech*, certains réalisateurs célèbres s'associent également à des modèles de génération : James Cameron rejoint le *board of directors* de *Stability AI* derrière *Stable Diffusion* en 2024¹, Darren Aronofsky se lance, avec son studio *Primordial Soup*, dans un partenariat avec *Google DeepMind* dont la première production à ce jour est le court métrage *Ancestra* d'Eliza McNitt en partie réalisé avec le modèle *Veo 3*. On remarque, avec ces cas de collaboration, un intérêt mutuel du monde du cinéma et des *start-ups* de l'intelligence artificielle, chaque partie semblant voir dans l'autre un morceau de son avenir. Enfin, après la pré-production et la production, la post production est également concernée par des changements majeurs. On imagine assez bien comment le *compositing* traditionnel des *VFXs* pourrait être augmenté, voire complètement remplacé, par des outils génératifs. Des modèles comme *Runway Gen-4* offrent déjà la possibilité de traiter des séquences de prise de vues réelles pour y incruster des effets. D'autres modèles se spécialisent expressément dans le

¹D'ailleurs, preuve que les frontières entre IA générative et cinéma sont perméables, le pdg actuel de *Stability AI* est l'ancien pdg de *Weta*, l'un des principaux studios spéciaux numériques derrière *Avatar* ou *Le seigneur des anneaux*

compositing. *Wonder Dynamics*, récemment racheté par *Autodesk*, éditeur d'une suite de logiciels quasiment hégémonique dans le domaine de la synthèse 3D, propose un panel d'outils à destination des *VFXs* : incrustation d'acteurs dans de nouveaux décors, intégration de personnages 3D remplaçant des acteurs, *motion capture* sans capteurs à partir d'une simple vidéo, *tracking* de caméra, etc.. *Beeble AI* offre, quant à lui, la possibilité de générer un relief à partir d'une *rush* tournée pour retravailler, après coup, son illumination. Du côté du son et du doublage, *ElevenLabs* permet de générer des voix et des bruitages hautement réalistes tandis que *Flawless* permet même de changer un dialogue (par exemple en le traduisant dans une nouvelle langue) en modifiant le mouvement des lèvres des acteurs pour conserver une synchronisation parfaite.

En parallèle de cet intérêt de la part de l'industrie établie, tout un nouvel écosystème de studios et de réalisateurs se revendiquant de l'ère de l'IA est en train de se mettre en place. Au rang des studios, on pourrait citer *Promise*, *Secret Level* — derrière la publicité de *Coca-Cola* de Noël 2024 qui avait fait polémique pour sa désincarnation supposée —, *Primordial Soup*, déjà cité, de Darren Aronofsky ou encore *EDGLRD* fondé par le réalisateur Harmony Korine. Tous ces studios revendiquent l'IA générative comme technique première de création par rapport à d'autres outils plus traditionnels. Et l'on pourrait également évoquer le cas des unités créatives formées aux seins des entreprises de la tech dans le but de créer des films démontrant les capacités des modèles : le *Creative Lab* de *Google DeepMind* ou encore *Runway Studio*. Du côté, certaines figures identifiées comme appartenant au microcosme de l'IA générative commencent à émerger : Dave Clark réalisateur et co-fondateur du studio *Promise*, le duo d'artistes Canadien Vallée Duhamel dont l'un des courts métrages a été produit par *Runway Studio* ou encore Paul Trillo, réalisateur de clips dont le très remarqué *The Hardest Part*, généré avec le modèle *Sora*, qui utilise les possibilités d'interpolation de l'IA et ses hallucinations pour mettre en scène l'amour d'un couple et sa séparation (« the hardest part ») à travers différentes scènes et moments de vies qui se succèdent à un rythme effréné et surtout de manière continue, sans coupe apparente. Et toute cette production artistique trouve un public dans les nouveaux festivals de film générés par intelligence artificielle, plus nombreux chaque année, à l'image de *l'AI Film Festival* lancé par *Runway* qui offre un bel aperçu de l'état de la création, ou dans les festivals plus anciens qui orientent une partie de leur programmation vers cette nouvelle technologie comme *Tribeca* ou *Ars Electronica*. Mais, la distribution de ce nouveau médium est encore balbutiante et les tensions parfois vives : je me souviens d'avoir assisté, au festival d'Annecy, à la projection d'un court métrage réalisé par IA que l'ensemble de la salle, constituée en grande partie d'un public de professionnels de l'animation,

avait hué. On retrouve ici, au cœur des troubles, la question de l'usage de l'IA générative et de sa créativité, réelle ou prétendue. Si de nombreux studios se tournent certainement vers ce medium par soucis d'économie et de réduction de main d'œuvre, se livrant dans bien des cas à un pillage de propriété intellectuelle, on peut également accorder à de nombreux artistes le crédit d'une ambition créative sincère et originale. L'essor des clips musicaux utilisant cette technique, à commencer par ceux de Paul Trillo, ou des tentatives plus expérimentales comme le long métrage *Aggro Drift* d'Harmony Korine, tourné en caméras thermiques et recouvert d'une surcouche de motifs générés par IA, en fournissent la preuve.

Les modèles de génération

Nous l'avons vu, ce ne sont pas les outils et les modèles qui manquent dans le champ de la génération d'images, et, plus largement, de l'IA générative. Chaque semaine, de nouveaux modèles fleurissent sur les plateformes, renversant à tour de rôle l'hégémonie des uns et des autres, quand ce ne sont pas simplement de nouvelles architectures, exposées dans des articles scientifiques, qui bouleversent le paradigme du moment. Derrière l'intelligence artificielle générative capable de produire des images mais aussi de vidéos, de textes ou encore de sons, se cache un ensemble de termes mystérieux aux contours souvent flous et mal définis pour le grand public : *machine learning*, *deep learning*, réseau de neurones, *GANs*, modèles de diffusion... À quoi renvoient ces termes ? Quelles sont leurs liens réciproques ? Et surtout, par rapport au sujet qui nous intéresse : quelle est leur relation avec la génération d'images ?

Tout commence, en tout cas pour l'IA générative, avec le *machine learning*, catégorie sans doute la plus importante au sein de la grande famille de l'intelligence artificielle, qui repose sur un principe simple ainsi décrit par le chercheur Michael Wooldridge dans son livre *A brief history of artificial intelligence*¹ : « disposer de programmes capables de calculer un résultat souhaité à partir d'une entrée donnée, sans avoir reçu de recette explicite sur la manière de procéder. »². Au lieu de fournir au programme un algorithme pré-codé, c'est-à-dire une suite explicite de tâches à exécuter les unes à la suite des autres, le *machine learning* se contente de donner au programme une base de données à partir de laquelle il sera en mesure d'apprendre, seul, l'algorithme à suivre pour résoudre un problème donné. Le *machine learning* est donc une forme d'intelligence artificielle

¹ WOOLDRIDGE Michael J., *A brief history of artificial intelligence: what it is, where we are, and where we are going*, First U.S. edition., New York, Flatiron Books, 2021, 262 p.

² *Ibid.*, p. 95., Traduction personnelle : « to have programs that can compute a desired output from a given input, without being given an explicit recipe for how to do this. »

particulièrement utile pour résoudre des problèmes qui ne peuvent pas être résolus en suivant une simple recette. C'est le cas, par exemple, de la reconnaissance du texte manuscrit : il n'y a pas d'algorithme unique applicable de manière universelle car chaque lettre diffère de toutes les autres. Dans sa forme la plus classique, l'apprentissage du *machine learning* repose sur le *supervised learning* : le modèle apprend sans règles mais à partir d'une base de données classifiées. C'est le cas de notre exemple de la reconnaissance du texte manuscrit : le modèle a accès pendant son entraînement à une base de données recensant des millions d'exemples de lettres manuscrites labellisées. Tout l'enjeu de cet entraînement est de permettre au modèle d'apprendre la structure visuelle de chaque lettre de manière à pouvoir ensuite labelliser de manière autonome de nouveaux exemples non inclus dans le *dataset*. Le *machine learning* est un champ de recherche aussi vieux que l'intelligence artificielle mais dont le succès récent repose sur les *neural networks*, ou réseaux de neurones, qui imitent le fonctionnement des neurones biologiques transmettant un message binaire (activé ou éteint) selon le message qu'ils ont reçu au préalable d'un autre neurone :

Les réseaux neuronaux, comme leur nom l'indique, s'inspirent des réseaux de cellules nerveuses (les neurones) présents dans la microstructure du cerveau et du système nerveux. Les neurones sont des cellules qui peuvent communiquer entre elles de manière simple, grâce à des fibres cérébrales appelées axones, qui relient les neurones entre eux via des « jonctions » appelées synapses. [...] L'idée des réseaux neuronaux dans le machine learning consiste donc à utiliser ce type de structures dans un programme informatique ; après tout, le cerveau humain prouve que les structures neuronales peuvent apprendre très efficacement.¹

Wooldridge retrace l'histoire de la recherche portant sur les réseaux de neurones qui commence très tôt, dès les années 40, sous l'impulsion des cybernéticiens Warren McCulloch et Walter Pitts qui propose un premier modèle mathématique des neurones biologiques. Puis vient Frank Rosenblatt et son modèle de réseau neuronal, le *perceptron* inventé en 1957 et qui fait toujours sens aujourd'hui : c'est un neurone artificiel recevant des *inputs* auxquels correspond, à chaque fois, un poids différent, et fournissant un *output* binaire (*allumé* ou *éteint*) variant selon la valeur de la somme des poids de ses *inputs*. Si cette valeur est au-dessus d'un certain seuil d'activation, alors l'*output* est actif, sinon il reste passif. En suivant cette logique, plus le réseau de neurones dispose de couches, plus il est précis. Mais bien que cette idée d'organisation en couches soit présente dès le début de la recherche

¹ *Ibid.*, p. 99., Traduction personnelle : « Neural nets, as the name suggests, were inspired by the networks of nerve cells—neurons—that are found in the microstructure of the brain and nervous system. Neurons are cells that can communicate with each other in a simple way, through fibers in the brain called axons to other neurons via “junctions” called synapses. [...] The idea of neural nets in machine learning, then, is to use these kinds of structures in a computer program; after all, the human brain provides proof that neural structures can learn very effectively. »

chez McCulloch et Pitts, elle n'a pas pu être mise en pratique pour une raison très prosaïque : on ne savait pas comment entraîner un tel réseau sans même parler de la puissance de calcul nécessaire, largement hors de portée à l'époque. La recherche dans le champ des *neural nets* a donc été abandonnée à la fin des années 60 au profit de l'IA symbolique qui l'a emporté sur le moment. Mais, après plusieurs échecs, les réseaux de neurones reviennent à la charge vers 2005 avec un nouveau modèle : le *deep learning*, ou apprentissage profond. Celui-ci est « profond » pour trois raisons : il dispose de nombreuses couches, de nombreux neurones et de nombreuses connexions. Si bien que les réseaux de neurones actuels atteignent des proportions impressionnantes : environ 100 millions de neurones pour GPT-3, soit le même ordre de grandeur que le cerveau humain. Une question demeure avec les *deep neural networks* : comment les entraîner ? Geoff Hinton, l'un des chercheurs les plus importants dans le domaine du machine learning fournit une réponse à la fin des années 2000 : il faut à la fois de la *data* et de la puissance de calcul. La puissance de calcul est fournie par les *GPU*, *Graphic Processing Units*, qui permettent de mener en parallèle un grand nombre d'opérations et qui avaient déjà permis l'arrivée de la 3D temps réel. Concernant la *data*, des initiatives se développent pour constituer des bases de données assez importantes pour pouvoir entraîner des réseaux de neurones de manière efficace. C'est dans ce but qu'est lancé, en 2007 par la chercheuse Fei Fei Li, le projet *ImageNet*, bibliothèque de 14 millions d'images labellisées :

L'intuition de Li était que l'ensemble de la communauté du *deep learning* bénéficierait de disposer de grands ensembles de données bien entretenus qui fourniraient une base commune à partir de laquelle les nouveaux systèmes pourraient être formés, testés et comparés. À cette fin, elle a lancé le projet ImageNet.¹

Et de fait, l'intuition de Li se confirme avec le succès d'*AlexNet* en 2012, programme de reconnaissance d'image conçu par Geoff Hinton, Alex Krizhevsky et Ilya Sutskever, entraîné avec *ImageNet* et battant de loin tous les records de reconnaissance d'image instaurés par des modèles précédents. Mais ce succès du *deep learning* s'est aussi accompagné d'une mise en lumière de ses limitations. Wooldridge en distingue deux : d'une part leur opacité, les *neural nets* fonctionnant comme des boîtes noires dont il est impossible d'expliquer le raisonnement ; d'autre part, leur manque de robustesse, les réseaux de neurones pouvant donner des résultats différents pour des cas apparemment similaires. Par ailleurs, le *deep learning* s'était appuyé jusqu'à là sur de l'apprentissage supervisé dont on a évoqué la logique plus haut. Ce type d'entraînement fonctionne

¹ *Ibid.*, p. 104., Traduction personnelle : « Li's insight was that the entire deep learning community would benefit from having large, well-maintained data sets that would provide a common baseline against which new systems could be trained, tested, and compared. To this end, she started the ImageNet project. »

bien pour des tâches de classification lorsqu'on cherche à obtenir une réponse unique pour une entrée donnée. C'est le cas, notamment, de la reconnaissance d'images : n'importe quelle image de canard implique un label unique en sortie — « canard ». Mais le *supervised learning* rencontre une limite majeure dans le cas de la génération d'images : il ne permet pas de générer de *data* nouvelle. Si les modèles comme *AlexNet* sont très bons à la reconnaissance d'images — fournir un label textuel à partir d'une image — le processus inverse ne fonctionne pas : impossible de générer une image entièrement nouvelle à partir d'un label. En effet, l'apprentissage supervisé entraîne des modèles à fournir un *output* de très faible complexité, par exemple un label, à partir d'un *input* de très haute dimension, par exemple une image de plusieurs millions de paramètres. Or, dans le cas de la génération d'images, et de l'IA générative en générale, le modèle doit fournir un *output* hautement complexe, de même dimension que l'*input*. Le *generative machine learning*, sous-catégorie du *deep learning*, désignant les modèles capables de générer de la *data* semblable à celle du *set* d'entraînement mais néanmoins entièrement nouvelle, pose alors de nouveau le problème de l'apprentissage. Comment entraîner de tels modèles ?

En réaction aux limites du *supervised learning*, émerge un nouveau champ d'étude : l'*adversarial machine learning* qui consiste à mettre deux programmes en compétition, l'un cherchant à tromper l'autre. Cette technique aboutit au *reinforcement learning*, apprentissage par renforcement à partir de *feedback* d'erreur et de réussite. Google *DeepMind* développe plusieurs modèles dont l'entraînement repose sur ce principe dans les années 2010. D'abord un programme qui joue au jeu *Atari*, puis les fameux *AlphaGo* et *AlphaGo Zero* qui apprennent le jeu de Go. Après une première phase d'entraînement supervisé d'après un *dataset* de parties existantes, succède une phase d'entraînement par renforcement durant laquelle le programme s'améliore uniquement en jouant contre lui-même, sans s'appuyer sur une base de données. Cette méthode donne des résultats sans précédents pur *AlphaGo* et *AlphaGo Zero* dont l'originalité des coups, jamais rencontrés chez aucun joueur humain, est remarquée. Dans le champ de la génération d'images, l'*adversarial learning* constitue également une avancée majeure. Ian Goodfellow, chercheur chez Google travaillant sur des tâches de reconnaissance d'image, réfléchit à une manière d'inverser le processus : non pas aller de l'image au texte, mais du texte à l'image. Il trouve une solution en 2014 en inventant les *GAN*, *Generative Adversarial Network* qui reprend le principe de l'*adversarial learning* pour l'appliquer à la génération d'images : un réseau générateur est mis en compétition avec un réseau discriminateur ; le générateur produit des images, à partir de bruit aléatoire, qui sont

ensuite incluses à un *dataset* contenant également des images réelles ; ce *dataset* est alors présenté au discriminateur qui essaie de déterminer, pour chaque image présentée, si celle-ci est vraie ou fausse : chaque succès du discriminateur est un échec du générateur et inversement. Par ce processus d'aller-retour, générateur et discriminateur s'entraînent mutuellement et s'améliorent. Plus le *GAN* tourne, plus le générateur produit des images réalistes et plus le discriminateur progresse dans la détection de *fakes*. Lorsque le générateur parvient à tromper le discriminateur avec une certaine régularité — c'est-à-dire lorsque la prédiction du réseau discriminateur que la probabilité qu'une image soit fausse est constamment de 0,5 — l'entraînement est un succès. Les *GANs* offrent des résultats hautement convaincants, comme en témoigne le site thispersondoesnotexist.com qui s'appuie sur le modèle *StyleGan2* développé par *Nvidia*, et qui génère, de manière aléatoire, à chaque chargement, un portrait humain indistinguishable d'une photographie réelle. Mais, en dépit de leur grand réalisme, les *GANs* souffrent d'un manque de stabilité dans l'entraînement qui les conduit souvent à « halluciner », c'est-à-dire à produire des images aberrantes comme des mains à six doigts.

Peu à peu les *GANs*, perdent en popularité et sont remplacés par les *diffusion models*, ou modèles de diffusion. Ils constituent aujourd'hui les principaux modèles de génération d'image comme *Midjourney* ou *Stable Diffusion*. C'est l'ingénieur Jascha Sohl-Dickstein qui, le premier, imagine l'application du principe de la diffusion thermodynamique (décrivant la manière dont des particules se diffusent dans un espace physique) au *machine learning*. Dans son article de 2015, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*¹, Sohl-Dickstein expose son intuition :

L'idée fondamentale, inspirée de la physique statistique hors équilibre, consiste à détruire systématiquement et lentement la structure d'une distribution de données par un processus itératif de diffusion vers l'avant. Nous apprenons ensuite un processus de diffusion inverse qui restaure la structure des données, ce qui donne un modèle génératif des données très flexible et facile à manipuler.²

Très concrètement, dans le cas de la génération d'image il s'agit d'apprendre à un modèle à débruiter une image de bruit pure jusqu'à obtenir une image faisant sens, représentant quelque

¹ SOHL-DICKSTEIN Jascha, WEISS Eric, MAHESWARANATHAN Niru, et al., « Deep unsupervised learning using nonequilibrium thermodynamics », s.l., pmlr, 2015.

² *Ibid.*, Traduction personnelle : « The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. »

chose. Les modèles de diffusion reposent alors sur un double processus. D’abord, le *forward process*, qui correspond à la phase d’entraînement pendant laquelle le modèle apprend à débruiter une image de bruit. Pour ce faire, le modèle dispose d’un *dataset* d’images nettes auxquelles il rajoute du bruit par paliers progressifs. À chaque palier, il dispose alors de l’image nette et de sa version bruitée. C’est un cas de *self-supervised learning* : le modèle se construit son propre *dataset* à deux entrées (à chaque x correspond un y , en l’occurrence ici, pour chaque image bruitée correspond une image nette) lui permettant ainsi d’apprendre à quoi ressemble une même image selon différents niveaux de bruit. En multipliant le processus avec des millions, voire des milliards d’images, et des dizaines ou des centaines de niveaux et de variations de bruit, le modèle apprend finalement le « chemin » vers n’importe quelle « bonne » image, i.e. représentant véritablement quelque chose, à partir de n’importe quelle image de bruit. Vient alors le second processus, le *reverse process* qui correspond à la partie générative. Son apprentissage terminé, le modèle est en mesure de partir d’une image de bruit parfaitement aléatoire et de débruiter cette image jusqu’à obtenir une représentation valide. Le processus de débruitage diffère légèrement selon les modèles : certains comme *Stable Diffusion* cherchent à prédire le bruit qui a été rajouté pour le retirer, d’autres prédisent plutôt l’image à l’étape suivante, c’est-à-dire légèrement débruitée. Quoi qu’il en soit, la force des modèles de diffusion tient à leur processus itératif qui peut inclure des centaines d’étapes, permettant d’obtenir des images très fines et très réalistes. S’il est en effet difficile d’obtenir d’un seul coup un résultat probant à partir de bruit pur, lorsque l’on procède par petites touches successives les linéaments d’une image se dessinent peu à peu, les contours et les formes apparaissent de plus en plus clairement et finissent par s’imposer comme des évidences. Tel un sculpteur qui, à force de milliers de gestes adroits fait émerger une figure d’un bloc de marbre, les modèles de diffusion exhument des images d’un brouillard de *pixels*.

Il existe encore d’autres grandes familles de modèles de génération d’images, comme les *VAEs* et les *Transformer-based generative models*, mais ceux-ci sont moins courants et offrent des résultats moins probants. En revanche, ils peuvent être utilisés en conjonction avec des modèles de diffusion pour une efficacité et une rapidité augmentées. C’est le cas des *VAEs* dont l’architecture facilitant la manipulation de *data* est intégrée à une sous-catégorie des modèles de diffusion — les modèles de diffusion latente — dont font partie *Midjourney*, *Stable Diffusion*, mais aussi la plupart des modèles de *text-to-video* comme ceux de *Runway*. C’est que la génération de vidéo doit répondre à un défi de taille : la consistance temporelle. Il faut que les différentes *frames* qui constituent une séquence prolongent un mouvement unique de manière uniforme et crédible, dans la

plupart des cas en respectant les lois physiques. L'enjeu de la persistance temporelle n'est pas seulement technique, mais ontologique : ce qui fait un film, ce n'est pas une suite d'images, mais la continuité perçue. L'IA doit donc simuler, non seulement une image, mais un temps cohérent, c'est-à-dire une cinématographie en acte. Il n'est donc pas possible de traiter chaque *frame* comme s'il s'agissait d'une simple image indépendante qu'on pourrait générer avec un modèle de diffusion classique, sous peine d'obtenir des résultats instables, clignotants. Nous avons déjà évoqué ces premières tentatives de génération de vidéo qui avaient envahi *internet* autour de 2022 de séquences hallucinatoires interpolant sans cesse des figures sans parvenir à se fixer sur aucun objet. Ainsi, la plupart des modèles de vidéo actuels comme *Runway* ou *Veo 3*, reposent sur le principe de la diffusion mais, en plus de prendre en compte la dimension spatiale des images, ils doivent considérer également la dimension temporelle de la séquence. Celle-ci constitue une nouvelle dimension qui se rajoute aux deux dimensions des images. Mais cette consistance temporelle est d'autant plus difficile à rendre qu'elle implique, avec le rajout d'une nouvelle dimension, une quantité exponentielle de calculs, et que les bases de données de vidéos pour entraîner les modèles sont beaucoup moins riches que celles d'images fixes. Pour économiser de la puissance, des nombreux algorithmes commencent donc par générer des vidéos de faible résolution et de faible fréquence pour ensuite les raffiner petit à petit. *Imagen* de *Google* orchestre par exemple sept modules successifs de *TSR* (*Temporal Super Resolution*, pour encore augmenter la fréquence d'images) et *SSR* (*Spatial Super Resolution*, pour augmenter la résolution spatiale). Aujourd'hui, des modèles comme *Veo 3* on atteint une consistance temporelle telle qu'ils sont capables de produire des vidéos avec toutes sortes de simulations physiques, décrivant par exemple le comportement des fluides, avec une précision extrême. Avec la génération d'image, puis de vidéo, l'IA générative franchit des paliers successifs dans le photoréalisme : les images produites sont toujours plus proches d'images réelles sans pour autant n'avoir jamais été tournées. Ce petit miracle repose sur un principe commun à toutes les architectures de modèle : l'échantillonnage dans une distribution probabiliste d'images.

***Pixel space* et espace latent**

Le *deep learning* repose en grande partie sur les principes mathématiques des probabilités, et la génération d'images n'y échappe pas. Quelle que soit la méthode d'entraînement et le modèle (*GAN*, *VAE*, Diffusion), le principe est toujours le même : que le modèle comprenne la structure de

la distribution des images d'entraînement pour produire de nouveaux échantillons, i.e. de nouvelles images extraites de la fonction d'approximation de cette distribution. En mathématiques, une distribution de probabilités correspond à la distribution d'un ensemble d'événements possibles en fonction de leur fréquence. Il existe différents types de distributions ; l'un des plus courants, et celui qui correspond à la distribution d'images d'un *dataset* dans un espace probabiliste est la distribution gaussienne, ou normale, qui regroupe la plus grande partie des événements autour d'une valeur centrale. C'est le cas, par exemple, de la distribution des tailles d'individus humains : en prenant un groupe assez important, on s'aperçoit que la majeure partie des individus se rassemblent autour d'une taille moyenne de 1,70m, très peu d'individus mesurant moins de 1,50m et plus de 1,90m. La fréquence des événements est alors élevée autour de la valeur moyenne, tandis qu'elle s'estompe à mesure qu'on s'en éloigne, formant une courbe en cloche. Une distribution comporte autant de dimensions qu'il y a de valeurs pour chaque cas. La distribution de probabilités concernant la taille d'individus ou les résultats d'un lancer de dé est unidimensionnelle car il n'y a, à chaque fois, qu'une seule valeur pour chaque événement (une taille pour un individu donné, un chiffre de 1 à 6 pour un lancer). Pour une image, en revanche, il y a généralement des centaines de milliers voir des millions de dimensions — autant qu'il y a de *pixels*, ou plus précisément de canaux. Ainsi, une image réduite au strict minimum, par exemple une image de deux *pixels* en niveau de gris, comporte deux paramètres et donc deux valeurs par événement, i.e. par image. La distribution du groupe d'images de deux *pixels* en niveau de gris comporte alors deux dimensions. Mais plus la résolution de l'image augmente, plus cela se complexifie, d'autant plus si l'on passe du niveau de gris, qui n'implique qu'une seule valeur par *pixel*, au RGB 8 *bits*, ce qui multiplie par trois le nombre de paramètres. Concrètement, pour une image carrée de 1000 *pixels* de côté, et un format RGB 8 *bits*, donc comportant pour chaque *pixel* trois valeurs allant de 0 à 255 codant la couleur de ce *pixel*, il y a un million de *pixels* et 16 777 216 de couleurs possibles pour chacun d'entre eux, ce qui donne $16\,777\,216^{1\,000\,000}$ possibilités d'images. C'est un chiffre énorme, bien plus élevé que le nombre d'atomes dans l'univers. D'un point de vue mathématique, cet ensemble de possibilités peut être vu comme un hypercube de trois millions de dimensions (correspondant au trois millions de paramètres de l'image : un million de *pixels* multiplié par les trois canaux RGB) et de 256 unités de côté (correspondant au 256 valeurs possibles pour chaque paramètre en 8 *bits*). Toutes les images de ce format — toutes les configurations qui peuvent exister — sont contenues dans cet espace probabiliste de très haute dimension qu'on appelle le *pixel space* ou *image space*. Chaque image peut ainsi être considérée

comme un point dans cet espace immense, ou plutôt comme un vecteur de trois millions de dimensions. Tout l'enjeu de l'apprentissage des modèles de génération d'images consiste alors à apprendre à se déplacer dans cet espace vertigineux des possibles pour aller d'une image de bruit, où la couleur de chaque pixel est déterminée aléatoirement, à une « bonne » image, c'est-à-dire une image ayant un sens, représentant quelque chose. Pour ce faire, le modèle dispose d'un *dataset* gigantesque de bonnes images dont il doit structurer la distribution en concevant une fonction d'approximation, ce qui revient à créer une sorte de cartographie de la position des images du *dataset* dans l'espace des images possibles. À force d'affinage, cette cartographie se précise et permet au modèle de repérer des chemins — en l'occurrence des caractéristiques dans l'agencement de *pixels* — menant à des lieux dans le *pixel space* où la densité de bonnes images est plus élevée. Il est alors de plus en plus simple d'échantillonner de bonnes images, c'est-à-dire d'extraire des images structurées de la distribution de toutes les images possibles.

Mais la navigation dans le *pixel space* représente un défi de taille. Celui-ci n'est pas structuré : deux images apparemment identiques peuvent être très éloignées dans le *pixel space* si la valeur de quelques-uns de leurs *pixels* change drastiquement, et, entre elles, peuvent résider des milliards d'images de bruit pur ne ressemblant à rien. C'est que la très vaste majorité de l'espace des images possibles est vide, ou plutôt rempli d'images de bruit aléatoire, et les bonnes images représentant réellement quelque chose constituent de fragiles îlots isolés. Comment alors est-ce qu'un modèle apprend à échantillonner correctement dans cette distribution, autrement dit, comment apprend-t-il à se déplacer dans cet espace immense ? Ici, nous étudions le cas des modèles de diffusion qui sont les plus performants à cette tâche. On a vu que ceux-ci partent d'une image aléatoire pour générer une nouvelle image. Vu l'immensité de cet espace dont l'écrasante majorité est rempli d'images dénuées de sens, cette image aléatoire est une image de bruit — il faudrait une chance impensable pour tirer au hasard une image déjà structurée — qui se trouve dans un endroit quelconque au milieu du *pixel space*. Il s'agit alors de se déplacer, depuis ce point, vers des îlots d'images structurées. C'est tout l'enjeu de l'étape d'apprentissage du *forward process* : en partant d'une image nette et en la bruitant légèrement, le modèle s'éloigne de cette image dans le *pixel space*. Il dispose alors de la trajectoire entre l'image nette et sa version bruitée. En répétant ce processus un très grand nombre de fois avec différentes variations et niveaux de bruit, le modèle parcourt un grand nombre de trajectoires possibles pour atteindre l'image nette, structurée. On retrouve ici l'analogie de la cartographie approximative du *pixel space* (une véritable cartographie

exhaustive serait impossible) qui lui permet, depuis n'importe point de cet espace (donc depuis n'importe quelle image bruitée) de retrouver son chemin vers l'image originale. En répétant l'entraînement avec des milliards d'images, le modèle augmente sa cartographie de l'espace des images possibles, apprenant les milliards de chemins menant vers des milliards d'images. Le *U-net* (le réseau qui dénoise) agit ainsi comme une boussole pointant vers les zones de plus hautes densité dans la distribution des images du *dataset*. Il peut, dès lors, échantillonner de nouvelles images à proximité de ces points chauds où il est presque sûr de rencontrer de bonnes images. Le *pixel space* n'est donc pas structuré en lui-même, géométriquement, mais c'est le modèle qui l'organise en apprenant à distinguer les bonnes images des mauvaises. En effet, si au départ l'algorithme ne fait pas la différence, il découvre peu à peu les règles auxquelles obéissent toutes les bonnes images ; par exemple, deux *pixels* côte à côte doivent avoir une couleur très proche. La mise-à-jour de telles caractéristiques permet ensuite de modéliser des chemins de probabilités. Ainsi, un modèle de diffusion, une fois entraîné, est en mesure de prendre en entrée une image de bruit — un point aléatoire du *pixel space* — à partir de laquelle il se déplace, par itération, jusqu'à atteindre une « bonne image ». Chaque étape de débruitage constitue donc un vecteur de déplacement dans l'espace des images possibles. Mais comment expliquer alors qu'un modèle de diffusion transforme une image de bruit en chat, par exemple, et une autre en voiture ? En réalité, il n'y a ni choix arbitraire du modèle ni reconnaissance d'un signal caché dans l'image de bruit. Ici, c'est le hasard qui est maître : tout dépend du tirage au sort du motif de bruit donné en entrée et de sa position dans l'*image space*. Si celle-ci s'avère être plus proche des images de chats que des images de voiture, c'est une image de chat qui sera générée et inversement.

Mais, le *pixel space* étant immensément grand et non structuré géométriquement, même s'il est possible de le structurer de manière probabiliste en établissant une cartographie approximative, cette tâche est très lente et très coûteuse en calcul. Certains modèles de diffusion, parmi les premiers développés comme les *DDPM* (*Denoising Diffusion Probabilistic Model*), parviennent à travailler directement dans cet espace au prix d'une lenteur qui affecte leur fonctionnement, les rendant pratiquement inutilisables dans le cas de la génération de vidéo qui augmente de manière exponentielle la puissance de calcul requise. En réponse à ce problème, les modèles de diffusion latente, comme *Midjourney* ou *Stable Diffusion*, échantillonnent dans une approximation du *pixel space*, l'espace latent. Leur temps de génération en est drastiquement réduit, et leur utilisation facilitée, notamment dans le cas de la génération conditionnelle — i.e. lorsqu'on rajoute une

condition comme un *prompt* au lieu de s'en remettre au hasard. En effet, l'espace latent constitue une réduction importante du nombre de dimensions de l'espace dans lequel le *U-net* doit se repérer, passant de quelques millions à quelques centaines voire dizaines. C'est que les dimensions ne correspondent plus, ici, à l'ensemble des paramètres d'une image (les millions de pixels et leurs canaux) mais à des grandes familles de caractéristiques partagées par des groupes d'images. Dans un espace latent, « chat » ou « cercle » peuvent ainsi constituer des dimensions ; une image de chat aura alors, parmi ses paramètres, un vecteur « chat » d'une valeur très importante, tandis qu'il sera très faible pour une image de voiture qui, elle, en revanche, possèdera un vecteur cercle d'une valeur supérieure à celle de l'image de chat. L'espace latent est donc une compression du *pixel space*. Celle-ci est opérée par des *auto-encoders* qui, en identifiant et encodant des caractéristiques essentielles, réalisent un *embedding* : une représentation vectorielle dense d'un objet (texte, image, son, etc...) dans un espace de dimension réduite, construite pour préserver les relations sémantiques ou structurelles entre objets. Un *embedding* est donc un vecteur numérique, présent dans l'espace latent, qui encode le sens ou la structure d'un objet complexe. Plus cet *embedding* est dense, réduisant une image (ou tout autre objet) à un minimum de *features* essentielles, et plus l'espace latent est structuré, i.e. plus il est compact et moins il comporte de dimensions. Du fait de cette structure réelle, géométrique et non plus seulement probabiliste, les images semblables forment des *clusters*, des groupes géographiques dans l'espace. Cette structuration sémantique facilite grandement la navigation du modèle puisque les groupes de bonnes images forment comme des points d'attraction, des gradients : il faut imaginer ces *clusters* comme des pics se détachant d'un paysage plat ; le modèle peut alors « suivre la pente » pour s'orienter vers de bonnes images. Et cette navigation facilitée s'accorde particulièrement bien avec l'injection de *prompt* — une indication textuelle. Nous avons vu qu'une image était générée aléatoirement à partir de la position dans le *pixel space* ou l'*espace latent* de l'image de bruit qui sert de base à la génération. L'image générée correspond ainsi à l'image la plus proche de l'image de bruit. Mais cela n'est vrai que pour la génération inconditionnelle. Dès lors qu'on entre un *prompt*, ou tout autre condition, le modèle s'efforce de suivre l'injonction. Comment cela fonctionne-t-il ? Pour injecter un *prompt*, il faut, tout comme pour une image, le retranscrire sous forme de vecteur, c'est-à-dire qu'il faut réaliser un *embedding* du prompt : le condenser en un concentré sémantique. Dans le cas d'une image, on a vu que l'*embedding* consistait à repérer les caractéristiques visuelles essentielles — les motifs principaux formés par la grille de pixels. Dans le cas des mots, il faut isoler les caractéristiques

sémantiques essentielles, les unités premières de sens qui sont l'équivalent des motifs visuels. Tout l'enjeu du *word embedding* réside donc dans la représentation d'un mot comme un vecteur qui fait sens. Pour les images, c'est plutôt simple simple — en tant que grilles de *pixels* ce sont déjà des suites de chiffres — mais comment représenter des mots de manière numérique ? Qu'est ce qui rend deux mots similaires d'un point de vue sémantique et deux autres différents ? Doit-on utiliser la proximité orthographique ? Une syntaxe semblable ne garantit en rien une similarité de sens. Devrait-on alors s'en remettre à la définition des mots ? Mais un même mot peut comporter différentes définitions selon le contexte. Et justement, il semble que ce soit avec le contexte que l'on tienne notre meilleure piste. Comme l'écrit Wittgenstein : « La signification d'un mot est son emploi dans le langage »¹. Ainsi, deux mots sont similaires s'ils sont souvent utilisés dans des contextes similaires, et d'autant plus différents que leur contexte d'emploi diffère. Avec une représentation vectorielle, on peut alors mesurer la similarité de deux mots en fonction de la proximité des vecteurs qui les représente dans l'espace sémantique. C'est le principe du *contrastive learning* : en apprenant à reconnaître, non pas des labels, mais des structures, le modèle est en mesure de créer des groupes de caractéristiques contrastant avec d'autres groupes. C'est en suivant ce principe de contraste que l'espace latent, qu'il concerne les images ou les mots, se structure : les propriétés sémantiques semblables s'agglutinent pour former des *clusters*, des points chauds dans l'espace, plus ou moins proches les uns des autres en fonction de leur proximité sémantique. Dans le cas du texte, le modèle peut se déplacer dans l'espace latent textuel de la même manière qu'il se déplace dans l'espace latent des images : en effectuant des opérations vectorielles. Ainsi, partant du vecteur « King », lui retranchant le vecteur « Homme » et puis rajoutant le vecteur « femme » on aboutit à un nouveau mot qui fait sens : « Queen ». Une question se pose alors : comment superposer l'espace latent des images et du texte de façon que les *clusters* correspondants soient positionnés aux mêmes endroits ? On veut, par exemple, que le *cluster* correspondant aux images de chats occupe le même endroit de l'espace latent que le cluster sémantique des mots reliés au concept « chat ». C'est le rôle des algorithmes comme *CLIP* (*Contrastive Language Image Pairs*) utilisés par les modèles de génération que de produire cette superposition. Pour ce faire, on entraîne les modèles de diffusion, non pas seulement avec des images, mais également avec un ensemble de descriptions textuelles associées à ces images. Ainsi, en même temps qu'il apprend à débruiter des images, le modèle apprend aussi à associer ces opérations à des descriptions textuelles. Le texte

¹ WITTGENSTEIN Ludwig, *Recherches philosophiques*, Paris, Gallimard, 2014, p. 45.

devient alors un paramètre pris en compte dans le tracé du chemin menant d'une image de bruit à une image nette : il est dès lors possible d'influencer le débruitage avec un vecteur textuel. Ce vecteur est ainsi rajouté au vecteur qui déplace l'image bruitée à chaque itération de débruitage : le *prompt* agit comme une force directionnelle dans l'espace de génération. Il existe différentes méthodes pour réaliser ce genre d'injection textuelle. Le *Classifier-Free Guidance (CFG)*, utilisé par *Stable Diffusion*, permet, par exemple, de contrôler l'influence du *prompt* sur la génération. À l'entraînement, le modèle s'exerce sur deux *datasets* distincts : l'un constitué uniquement d'images, et l'autre d'images accompagnées de descriptions textuelles. Au moment de la génération, il est alors possible de combiner les deux trajectoires des deux variantes du modèle pour moduler le *guidance scale*, le degré de contrainte du *prompt*. Plus le *guidance scale* est faible, plus la condition est faible et plus l'image est générée de manière aléatoire, plus il est fort et plus le vecteur textuel pèse dans la trajectoire. Et, au-delà du texte, il existe d'autres modes de conditionnement comme l'*image-to-image*, presque aussi répandu que le *prompt*. Encore une fois, il existe plusieurs méthodes pour qu'une image puisse exercer une influence sur une génération. Soit cette image est encodée comme un vecteur sémantique qui sera ajouté à chaque étape de débruitage. Soit elle se substitue à l'image de bruit qui sert de base à toute nouvelle génération ; le modèle altère alors plus ou moins l'information de cette image de référence — il la bruite — selon que l'on désire rester au plus près de sa structure ou s'en éloigner.

En fin de compte, là où les modèles de reconnaissance apprenaient à associer des labels aux images, les modèles de diffusion apprennent à leur associer des directions dans le *pixel space* ou dans l'espace latent ; souvent, en tenant compte d'une ou plusieurs conditions qui peuvent être d'ordre textuel ou visuel. Comment cela se traduit-il dans l'expérience des créateurs ?

Des images sans monde

Au moment de l'écriture, aucun des modèles importants de génération de vidéo comme *Runway Gen-4*, *Sora*, ou *Veo 3* ne propose de représentation d'un espace tridimensionnel dans lequel une caméra virtuelle pourrait se mouvoir librement. Concrètement, on ne peut pas se déplacer dans l'image avec sa souris, ou tout autre dispositif interactif, comme on peut le faire dans un viewport 3D. En ce sens, la génération d'images par IA est très différente de la prise de vues réelles ou de la 3D. Si certaines options, comme le contrôle très limité d'un semblant de caméra dans *Runway*, tendent vers une plus grande immersivité, force est de constater que l'IA générative

ne permet pas, pour le moment, d'expérimenter un espace géométrique consistant. Cela tient au fonctionnement probabiliste des modèles de diffusion qui apprennent statistiquement comment les images changent lorsqu'on simule un mouvement, mais sans construire de géométrie explicite. Dès lors, les pseudo-mouvements de caméra sont en fait des variations dans l'espace latent qui, comme nous venons de le décrire, est un espace statistique de relations entre images et non une scène géométrique persistante ou semi-persistante au sens de ce que met en œuvre un moteur de rendu 3D. Ainsi, lorsqu'on parle de « caméra » pour désigner l'angle de vue d'une image générée, il s'agit là d'un emploi purement métaphorique. Il n'y a en réalité aucune entité correspondante située dans un espace géométrique à l'instar d'une caméra virtuelle dans *Blender* ou *Unreal Engine*. Ici, la « caméra » est plutôt un modulateur de génération : il s'agit, pour chaque *frame* d'une séquence vidéo, d'échantillonner une image de façon à entretenir l'illusion d'une continuité spatio-temporelle qui n'est pas toujours garantie. La continuité visuelle, bien qu'elle soit de mieux en mieux maîtrisée, est toujours le résultat du *prompting* ou de techniques comme l'interpolation de *frames*, mais ne repose pas sur un monde consistant simulé. Et le risque d'un manque de cohérence s'amplifie drastiquement d'une génération à l'autre. À chaque fois qu'on lance un nouveau processus de diffusion, on obtient une nouvelle image impossible à prédire. Ce manque de consistance est particulièrement gênant lorsque l'on veut réaliser un montage, générer différentes échelles de plans et angles de vue dans une scène pour reconstituer une continuité spatiale : on entre alors dans une lutte avec le modèle pour préserver un semblant d'unité dans l'apparence des décors, des costumes ou des personnages. Mais c'est peine perdue : l'IA générative constitue une illustration technologique de l'idéalisme métaphysique, générant des apparences d'espaces plus proches du rêve ou de l'hallucination que du monde physique. D'ailleurs, le terme d'« hallucination » est employé par les chercheurs pour désigner les aberrations comme les fameuses mains à six doigts de *Midjourney*. L'IA générative produit ainsi une forme d'idéalisme discontinu, sans unité d'apparition.

Cette double absence d'espace et de persistance rapproche l'expérience de réalisation par IA générative des processus créatifs reposant sur l'imaginaire. Les réalisateurs qui s'emparent de ce médium doivent apprendre à réaliser leur film dans un espace purement mental. Ils n'ont d'autre choix que de concevoir leur plateau de tournage a priori, sans autres possibilités d'exploration que les ressources de leur imagination. En cela, l'expérience de la génération de films est assez proche de l'animation 2D : le hors champ n'existe pas et, comme un dessinateur, il faut concevoir

mentalement les cadrages et la continuité spatiale d'après une représentation imaginaire. Avec une différence de taille tout de même : dans le cas de la génération par IA, les artistes doivent s'associer avec une machine, lui déléguer une part de contrôle, plus ou moins importante selon les outils à disposition. Ce processus d'interaction avec la machine nécessite alors des qualités particulières comme le *prompt engineering*, véritable science du *prompt*. Il faut connaître les modèles, s'engager dans une véritable recherche pour être en mesure d'écrire les descriptions les mieux adaptées, selon l'algorithme, pour obtenir les meilleurs résultats. Ces *prompts* peuvent parfois être particulièrement longs et détaillés, à l'instar de cet exemple ayant servi à générer, avec le modèle *Veo 3*, un clip, impressionnant de réalisme, d'un véhicule roulant dans une flaque de boue et éclaboussant la lentille de la caméra :

La scène explose avec l'énergie brute, viscérale et imprévisible d'un rallye tout-terrain hardcore, capturée avec une esthétique dynamique, presque comme un *found-footage* ou un documentaire sportif embarqué. La caméra est souvent tremblante, semblant être montée à l'intérieur d'un des véhicules ou tenue par un spectateur audacieux très proche de l'action, fréquemment éclaboussée de boue ou d'eau, capturant des reflets involontaires de la lumière naturelle, souvent crue, du soleil filtrant à travers les arbres ou se reflétant sur les surfaces humides. Nous sommes plongés dans un environnement naturel hostile et sauvage, peut-être un sentier forestier dense et boueux, une pente rocheuse traîtresse jonchée d'éboulis instables, ou une série de traversées de rivières peu profondes et au courant rapide. Plusieurs véhicules tout-terrain fortement modifiés, totalement méconnaissables et sans marque, sont engagés dans une course effrénée et sans limites. Ce ne sont pas des modèles de salon, mais des machines robustes construites sur mesure : des buggies à roues ouvertes avec des moteurs exposés et des arceaux de sécurité proéminents, des camionnettes lourdement blindées avec des pneus surdimensionnés et crantés et des échappements à snorkel, dont la forme d'origine et le constructeur sont complètement masqués par des modifications extrêmes, des couches de boue incrustée et une impression générale de fonctionnalité brutale. Les sons dominants sont le rugissement assourdissant et guttural des moteurs puissants et indomptés, le sifflement des transmissions, l'impact brutal de la suspension qui touche le fond et les projections constantes de boue et d'eau. En l'espace de 8 secondes, l'un des véhicules de tête, un buggy surbaissé à cockpit ouvert, tellement recouvert d'une épaisse boue brune que sa couleur d'origine est une énigme, s'approche à une vitesse incroyable d'un large gué peu profond. Sans la moindre hésitation, son conducteur invisible s'engouffre directement dans l'eau. L'impact projette dans les airs une énorme nappe d'eau boueuse, presque solide et opaque, mélangée à des pierres et des débris provenant du lit de la rivière, qui engloutit complètement le petit buggy pendant un moment terrifiant, le dissimulant à la vue comme s'il avait été avalé par la rivière elle-même. Juste derrière lui, un camion personnalisé, tout aussi recouvert de boue, arrive au gué au moment où cet énorme mur d'eau atteint son apogée. Il s'agit d'un mastodonte doté d'une cage de sécurité externe fortement renforcée et sans aucun logo identifiable. Au lieu de ralentir ou d'essayer de trouver un chemin plus dégagé, le conducteur du camion, avec une agressivité absolue, plonge directement à pleine vitesse dans ce rideau opaque et tumultueux d'éclaboussures boueuses. Une fraction de seconde plus tard, le camion

surgit de l'autre côté du déluge, l'eau ruisselant de son toit et de son châssis, ses essuie-glaces surdimensionnés s'efforçant frénétiquement de dégager le torrent d'eau boueuse qui obscurcit la vision du conducteur. Il atterrit lourdement sur la rive opposée, la suspension grinçant, mais toujours à la poursuite du buggy qui réapparaît. Cette séquence palpitante, chaotique et visuellement spectaculaire, dans laquelle un véhicule crée un obstacle environnemental massif et le suivant le surmonte immédiatement par la force brute, constitue l'événement central, immersif et captivant de cette séquence de 8 secondes. La course se poursuit avec une férocité intacte, le terrain naturel lui-même participant activement au conflit.¹

Le niveau de description est impressionnant pour une séquence de seulement huit secondes. C'est à ce prix qu'est atteint un photoréalisme confondant et la précision du *prompt engineer*. Dans le cas du *text-to-video*, la réalisation de film semble étrangement se rapprocher de l'écriture, le réalisateur se change en un écrivain qui doit savoir manier l'hypotypose à la perfection. La mise en scène devient ici une sorte de « mise en prompt », de retranscription textuelle de la scène qu'on imagine. Le processus d'écriture est alors ramené au cœur du processus créatif. Là où sur un film classique, qu'il soit d'animation ou de prise de vues réelles, le scénario constitue un premier brouillon avec lequel on prend rapidement ses distances au moment de réalisation — souvent le scénario se heurte à un mur de réalité, telle ou telle scène qui fonctionne bien à l'écrit se révèle irréalisable en pratique pour des problèmes techniques, physiques, financiers, etc... — avec le *text-to-video*, le scénario reste d'actualité et même plus, il est intégré, partie prenante de la réalisation des images. Le texte entretient un lien direct, pas seulement symbolique, mais pratique avec sa manifestation visuelle : il est performatif, il donne naissance à l'image qui l'actualise. Cette performativité du texte est particulièrement visible dans l'inclusion de tout un vocabulaire cinématographique et photographique : *close-up*, *travelling*, *pan*, téléobjectif, 35mm, *fisheye*... Le réalisateur Dave Clark, dans une discussion organisée par le festival de Tribeca autour du modèle *Veo* ³², raconte comment cette performativité du texte lui a permis de vivre pleinement son double statut de scénariste-réalisateur : d'ordinaire, on demande toujours au scénariste de respecter une séparation nette avec la réalisation en n'incluant aucune indication de mise en scène dans le scénario ; avec la génération d'image, cette séparation n'a plus lieu d'être : le moment de l'écriture est aussi le moment de la réalisation. Il faut produire une écriture visuelle, décrire dans le détail les inflexions de la caméra, les échelles de plans ou encore les intentions lumineuses. En l'occurrence, dans l'exemple de

¹ On peut retrouver le *prompt* dans sa version originale et visionner la génération à laquelle il a donné lieu sur la page d'accueil du modèle Veo : <https://deepmind.google/models/veo/>

² On peut visionner la discussion sur la chaîne Youtube de Tribeca : <https://youtu.be/1M6RqtYEcA0?si=YoP2czxiGHdvDecs>

prompt que nous avons fourni, on peut remarquer les nombreuses références aux mouvements de la caméra (« souvent tremblante, semblant être embarquée à l'intérieur d'un des véhicules ou tenue par un spectateur audacieux très proche de l'action »), la mention d'artefacts avec les *lens flares*, le style d'image inspiré de l'esthétique du *found footage* ou des documentaires de sport, jusqu'aux éclaboussures sur la lentille. Encore une fois, comme dans le cas de la 3D, le réalisme des images est indexé au photoréalisme : la ressemblance, non pas avec la réalité telle qu'on la voit, mais avec la réalité telle qu'elle est enregistrée par un appareil photographique ou une caméra. Et, comme pour la 3D où l'on recrée de manière absolument artificielle des artefacts qui n'ont pas lieu d'être avec une caméra virtuelle, on ne peut s'empêcher d'apprécier, avec notre œil conditionné aux images photographiques, le niveau suprême de réalisme apporté par les éclaboussures qui coulent sur la lentille d'une pseudo-caméra dont nous sommes parfaitement convaincus de l'existence. Pourtant, à la différence de la 3D, il n'y a même pas de caméra, au moins virtuelle ; cette pseudo-caméra n'est plus que le fantôme, pourtant bien présent, d'une culture visuelle qui imprègne nos cerveaux mais aussi les bases de données (constituées en majeure partie d'images photographiques) qui servent à entraîner les modèles génératifs. Il semble que nous ne soyons pas près d'échapper au règne de l'image photographique.

Mais on a vu que le *prompt* n'est pas le seul moyen de contrôle de la génération d'images. En réalité, d'autres modes de conditionnement comme l'*image-to-image* sont quasiment systématiquement utilisés en conjonction du *prompt* pour offrir plus de contrôle, principalement concernant la composition des plans. Car, réaliser un film c'est produire une continuité spatiale, bien souvent par le montage et la grammaire cinématographique. Or, la description textuelle, aussi précise soit-elle ne suffit pas toujours pour obtenir la composition idéale conçue de manière imaginaire dans un montage mental. Quand bien même on serait un virtuose de la description — et en cela de nombreux prompteurs font appel à des *LLMs* comme *ChatGpt* pour enrichir leurs *prompts* —, la plus détaillée d'entre elles est toujours terriblement limitée face aux richesses visuelles. Comme l'illustre l'adage populaire « une image vaut mille mots », il est impossible de décrire à la perfection une image, et le texte se révèle finalement un outil bien frustrant lorsqu'on ne possède que lui. Le processus de génération d'un ensemble de plans cohérent, pour monter une continuité spatiale, peut s'avérer un véritable combat contre le modèle quand il ne génère pas, malgré des dizaines d'itérations et de *prompts* plus précis les uns que les autres, d'image à la composition appropriée. C'est là qu'interviennent les modes de contrôle supplémentaire comme

l'*image-to-video* ou le *video-to-video*. Comme leurs noms l'indiquent, ces différents conditionnements permettent d'inclure, en plus d'un *prompt*, une image ou une vidéo comme référence de base. Ces outils s'avèrent précieux et sont très largement utilisés. L'*image-to-video* permet, par exemple, de fonctionner par étapes successives : on génère d'abord des images, qu'on sélectionne ensuite pour servir de base à la génération d'une vidéo qui conservera la même structure, le même cadrage et la même échelle de plan. Dans le cas où l'on fournit au modèle la première et la dernière *frame* d'une séquence, cela revient à utiliser l'algorithme comme un animateur réalisant une interpolation entre deux images clés. Et pour encore plus de contrôle, il est possible de donner en *input* des dessins de *storyboard* afin de guider la composition. S'il est difficile d'obtenir l'angle idéal en générant des images avec du simple texte (parfois cela revient juste à gagner au loto), il est plus simple de se tourner vers les méthodes traditionnelles comme le dessin qui offre un contrôle total. De nouveaux outils, comme ceux de *Runway Gen-4*, permettent alors d'utiliser ces croquis en tant que référence de composition. Avec cette manière de fonctionner, le *storyboard* peut acquérir une place centrale en jouant, tout comme le scénario, un rôle performatif. Et, au delà de la composition, l'image en *input* peut aussi influencer le style visuel de la séquence générée — c'est utile pour produire des vidéos « dans le style de... ». Enfin, le conditionnement *video-to-video* offre un contrôle encore accru par rapport à l'*image-to-video*. Cette fois-ci, c'est une séquence animée dont la structure, et particulièrement celle du mouvement, sert de base au modèle qui agit comme une sorte de moteur de rendu réalisant un traitement esthétique. On peut ainsi réaliser des effets spéciaux d'une grande complexité avec une facilité déconcertante : à partir d'une simple vidéo où l'on traverse notre salon, on peut, selon le *prompt*, se retrouver au milieu d'un désert lunaire ou marchant sur l'eau, portant un costume surréaliste ou même changé en créature fantastique.

Les modèles tendent vers une amélioration constante des possibilités de contrôle, condition pour que l'IA générative devienne véritablement utilisable dans des productions professionnelles. Parmi les modèles disponibles, ceux en *open source* offrent généralement les meilleurs outils de customisation. C'est le cas de l'interface *Comfy UI*, d'abord développée autour de *Stable Diffusion*, qui met à disposition des utilisateurs un système nodal — des blocs pouvant être connectés les uns aux autres — permettant de configurer une pléthore de paramètres pour de nombreux modèles. Il est, par exemple, possible de régler le nombre d'étapes de débruitage, de gérer l'influence du conditionnement (*prompt* ou image), de rajouter un *negative prompt* (ce qui ne doit surtout pas

apparaître dans la génération), de faire du *Style transfer*, du *Face swap*, d'extraire le squelette animé d'un danseur, de générer une *depth map* pour obtenir la composition 3D d'une image, etc... Les possibilités sont infinies et dépendent principalement de la créativité des utilisateurs dans les *workflows*, l'agencement des *nodes* entre eux. Cela revient à une forme de *compositing* augmenté, plus interactif parce qu'on utilise les capacités génératives de l'IA, mais toujours cantonné à l'espace bidimensionnel de l'écran qui reste une interface. Puisque le réalisateur travaille dans la surface de l'image, et non dans un volume, l'expérience de création est, on l'a vu, d'ordre imaginaire et graphique, plus que perceptive ou spatiale. Tout comme l'animateur 2D, il s'agit de composer le film directement en deux dimensions dans l'espace du cadre. On génère un grand nombre d'images, on les découpe et on les colle entre elles jusqu'à obtenir la composition adéquate pour générer la bonne vidéo. Les *workflows* de création de film avec de l'IA générative imposent d'ailleurs toujours de nombreux allers retours avec des logiciels traditionnels de *compositing* comme *Photoshop*. Ce régime d'expérience artistique repose principalement sur l'écriture, qui est réactualisée d'une manière très originale et inclue au cœur du processus de création, et sur les images qui constituent souvent une première base pour générer une séquence vidéo. Mais, malgré les possibilités créatives et la richesse des *workflows* existants, cette expérience de création reste désincarnée : elle se fait sans perception corporelle d'un espace. Le corps est le grand absent de l'IA générative. On retrouve ici la faible aura phénoménologique dont parlait Stéphane Vial à propos de l'écran numérique qui n'est pas un lieu à habiter mais seulement une surface à regarder. Ce n'est pas un monde au sens des critères de Merleau-Ponty évoqués précédemment : les interfaces des modèles de diffusions et les images qu'elles affichent ne constituent ni le lieu d'une intersubjectivité, ni une totalité, ni même l'horizon de notre perception. Dès lors, puisqu'elle ne permet pas d'immersivité, la génération d'images ne permet pas non plus le déploiement d'un point de vue ; elle ne saurait être une perception instrumentée : ce n'est pas une prolongation de notre subjectivité dans un monde virtuel car il n'y a simplement pas de monde. Cette absence de persistance spatiale et ontologique rend alors difficile l'application du principe de prise de vues virtuelles à la génération de vidéo. Contrairement à l'expérience d'un utilisateur de moteur 3D temps réel, aucun « prompteur » n'est en mesure, d'abord, de se déplacer dans un espace tridimensionnel en contrôlant son point de vue, et ensuite, quand bien même il le pourrait, de garantir une forme de consistance : avec l'IA générative, l'opérateur qui serait en mesure de cadrer un personnage dans un décors, s'il leur tournait le dos, ne serait pas assuré de les retrouver en pointant de nouveau, l'instant d'après, la

caméra sur eux. La génération d'image par IA est à cet égard davantage une forme d'animation automatisée qu'un tournage dans un monde. En revanche, si elle battue par la 3D temps réel pour ce qui est de l'engagement corporel, la génération de vidéo présente un avantage de taille : la création continue d'une imprévisible nouveauté. L'IA générative, et c'est le principe même de la génération, se donne sans cesse dans un surgissement d'images libre et sans entrave. En cela, elle remplit mieux que n'importe quel autre système de synthèse visuelle le quatrième critère phénoménologique du monde : l'inachèvement entraînant un apparaître constant.

Plasticité artificielle

Nous l'avons évoqué, l'IA générative est au cœur de tensions importantes dans le monde de l'art, et du cinéma en particulier. Depuis les grèves d'Hollywood jusqu'à la séance de projection à Annecy d'un clip réalisé avec *Stable Diffusion* qui est hué par la salle entière, en passant par le harcèlement en ligne qu'ont subi de nombreux artistes, l'IA générative excite les esprits et ne laisse personne indifférent. Les critiques sont nombreuses et plus ou moins fondées. Parmi celles qui nous semblent les plus légitimes, il y a bien sûr la dénonciation de l'usage purement financier que certains gros studios veulent faire de l'IA : réduire les coûts, faire du *dumping* social, sans compter la question du droit d'auteur ou du droit à l'image. Mais au-delà des thématiques socio-économiques, une critique bien plus fondamentale, bien plus profonde, porte sur la nature même de l'intelligence artificielle et la remise en cause de sa générativité. En effet, dans les commentaires souvent très agressifs que reçoivent par milliers les films générés par IA sur les réseaux sociaux, revient souvent l'accusation d'un manque de créativité, de non-originalité. Les modèles se contenteraient de copier tel quel le travail d'artistes humains : un pur recyclage de *pixels*. Plus que ça, l'IA générative serait une authentique menace pour la créativité humaine, remplaçant le génie spontané de l'humain par l'intelligence froide et calculatrice des machines. Cet argumentaire présente les accents d'un débat ancien, ouvert à la fin du XIX^e siècle avec les premières tentatives de mesure scientifique de l'intelligence, bien avant l'arrivée de l'intelligence artificielle donc. Dans son ouvrage *Les métamorphoses de l'intelligence*¹, la philosophe Catherine Malabou revient sur cet affrontement qui oppose « vie biologique » (quantitative, déterminé, naturelle, domaine de la science) et « vie symbolique » (qualitative, libre, humaine, domaine des sciences humaines), champ

¹ C. Malabou, *Métamorphoses de l'intelligence*, op. cit.

de bataille au milieu duquel se trouve le concept d'intelligence qui concentre les plus vives attaques comme les plus féroces défenses. Parmi les opposants à la notion d'« intelligence » renvoyée du côté de la vie biologique, bête et méchante, on compte de nombreux philosophes, au premier rang desquels Bergson. Pour ce dernier, l'intelligence est une « faculté d'adaptation », une sorte d'outil rigide caractérisé par « une incompréhension naturelle de la vie ». Ainsi, selon la Bergson, l'intelligence se trouve contre la vie et ne peut permettre la création, « cette évolution créatrice qui est la vie »¹. L'intelligence n'a alors d'autre choix que de s'aider de l'intuition pour sympathiser avec la vie : « À l'intelligence, l'intuition donne l'esprit, c'est-à-dire l'être qu'elle n'a pas. »². À leur tour, de nombreux penseurs et écrivains français comme Proust, Valéry ou Deleuze partagent les vues de Bergson, méprisant une intelligence froide, passive et mécanique, incapable de penser la création imprévisible qu'est la vie. Il y a donc, dans une part importante de la philosophie du XXe siècle, une opposition à l'intelligence dont l'espace est laissé libre pour d'autres disciplines comme la psychologie ou la biologie. Cette opposition féroce s'explique par la peur de « l'intelligence des psychologues et sa menace de normalisation, de standardisation et d'instrumentalisation de la pensée et des conduites. »³. Cette opposition de la philosophie à la notion intelligence, conçue comme résistance à un plan de soumission de la vie à la calculabilité machinique, semble converger exactement avec les positions, dans le monde de l'art, des détracteurs de l'IA générative qui lui refusent toute possibilité créative. L'intelligence artificielle, pas plus, et sans doute encore moins, que l'intelligence biologique n'est capable d'une véritable créativité, d'autre chose que d'un mécanisme stupide destiné à atrophier le génie humain. Or c'est justement ce génie qui est supposément au cœur de la création artistique. Dans *Artificial aesthetics*⁴, Lev Manovich revient sur ce « mythe de la créativité » selon lequel l'art est le domaine privilégié, voire exclusif, de la créativité humaine. Pourtant, cela n'a pas toujours été le cas et bien longtemps les artistes n'ont été que de simples artisans dont le talent tenait plus à un savoir-faire longuement appris qu'à une entité aussi mystérieuse et puissante que le génie. Les choses changent d'abord à la Renaissance et puis surtout au XIXe siècle avec le Romantisme qui exalte les passions irrationnelles et l'élan instinctif de l'artiste-créateur, faisant de lui la manifestation suprême de l'humain :

¹ H. Bergson, *L'évolution créatrice*, op. cit., p. 163.

² C. Malabou, *Métamorphoses de l'intelligence*, op. cit., p. 20.

³ *Ibid.*, p. 23.

⁴ MANOVICH Lev et ARIELLI Emanuele, *Artificial Aesthetics*, 2024., <https://manovich.net/index.php/projects/artificial-aesthetics>

Enfin, la vision moderne que nous considérons encore aujourd'hui comme acquise s'est imposée à l'époque romantique. Pendant plusieurs siècles, les artistes étaient simplement considérés comme des créateurs d'objets selon des règles établies, et l'art était dissocié des idées, de l'intellect et de l'éducation libérale. Aujourd'hui, cette tradition a été renversée. Un artiste romantique est supérieur à tous les autres. Il se situe entre Dieu et l'humanité. En fait, avec le déclin de l'influence de la religion à la fin du XIXe siècle et au début du XXe siècle, l'artiste a pris la place de Dieu.¹

Cette vision contemporaine de l'art est donc l'héritage d'une construction culturelle du XIXe siècle — Manovich parle d'une « invention de l'art ». D'une manière générale, l'héritage romantique instaure un primat du symbolique sur le biologique et le mécanique. Mais, comme le souligne Malabou, ce partage strict, s'il tenait encore jusqu'au milieu du XXe siècle, est sérieusement ébranlé voire rendu complètement impossible par la récente arrivée de l'IA, intriquant les trois aspects de l'intelligence :

La fragilité des frontières entre intelligence et intellect, cerveau et intellect, machine et intellect, intelligence (naturelle) et intelligence (artificielle), est devenue si manifeste qu'elle interdit tout partage assuré entre le biologique, le mécanique et le symbolique.²

À la suite de la découverte de la puce « neuro-synaptique » *TrueNorth* développée par *IBM* qui « fonctionne de facto comme un branchement synaptique. »³ permettant une plasticité artificielle, Malabou fait son mea culpa : après avoir défendu l'impossibilité de la plasticité artificielle, elle s'incline devant sa réalité en comparant ce processus d'acceptation à celui par lequel a dû passer le champ de la génétique avec la découverte de la plasticité cérébrale. L'intelligence artificielle, dans ses dernières avancées, est moins rigide qu'on aurait pu le croire. Tout comme les généticiens au cours du XXe siècle ont dû réviser leur conception du codage génétique pour admettre une certaine autonomie du phénotype (plasticité des comportements) vis-à-vis du génotype, il nous faut désormais accepter la capacité d'auto-configuration des modèles d'intelligence artificielle

¹ *Ibid.*, chap. 4, "AI and Myths of Creativity." Traduction personnelle : « Finally, the modern view that we still take for granted today was established during the Romantic period. For many centuries, artists were understood merely as makers of objects according to rules, and art was divorced from ideas, intellect, and liberal education. Now, this tradition has now been reversed. A Romantic artist is superior to everybody else. He is situated between God and humanity. In fact, as religion's influence waned throughout the later 19th and early 20th centuries, the artist took God's place. »

² C. Malabou, *Métamorphoses de l'intelligence*, op. cit., p. 23.

³ *Ibid.*, p. 108.

génération. Autrement dit, le code (génétique comme informatique) peut engendrer de l'indétermination ce qui ouvre la possibilité à une forme de créativité algorithmique :

Les algorithmes sont impliqués également dans des activités qualitatives, comme la création artistique, l'imitation d'un style, ou l'improvisation musicale. Tous les experts ont reconnu qu'AlphaGo par exemple, la machine qui a battu en mars 2016 Lee Sedol, un des meilleurs joueurs mondiaux de go, avait fait preuve d'une certaine créativité. C'est pourquoi l'opposition qu'établit le philosophe Bernard Stiegler entre le "calcul", le quantitatif, et l' "improbable", à savoir le qualitatif, n'est pas selon moi totalement pertinente. La subtilité du calcul algorithmique aujourd'hui tient à ce qu'il est précisément capable de simuler le non-calcul, c'est-à-dire la spontanéité.¹

La perte de contrôle que représente la plasticité artificielle, et qu'il s'agit d'accepter, correspond à ce que Malabou nomme la « quatrième blessure narcissique » après les trois premières blessures énoncées par Freud : « Après Darwin, après Copernic, après la psychanalyse, vient la quatrième blessure : la capture de l'intelligence par sa propre simulation. »². Au regard de la conception dominante de l'art que nous venons de décrire, selon laquelle l'artiste constitue la meilleure manifestation de l'exceptionnalisme humain, échappant à toutes les règles et toutes les mesures, on comprend que l'idée d'un l'art généré par intelligence artificielle fasse scandale. Difficile de renoncer à un tel privilège : l'investissement de ce sanctuaire par les machines est perçu par nombre d'entre nous comme un sacrilège. Mais, le traumatisme s'évanouira dès lors qu'on changera notre définition de l'automatisme, jusqu'à là perçu comme un déterminisme. Or, « comme dans un tableau de Pollock », l'automatisme n'est certes pas contingent, mais il n'est pas non plus condamné à la nécessité : « Le milieu entre nécessité et contingence est le véritable espace de vie des automatismes. »³. Dans le cas de la génération d'images, le processus d'échantillonnage dans l'espace immense des images possibles garantit presque à tous les coups la création d'une image entièrement nouvelle, c'est-à-dire d'un arrangement de *pixels* unique au monde. On le voit, il ne s'agit pas d'un simple recyclage, d'une recombinaison d'éléments d'images déjà existant. L'entraînement des modèles est justement conçu sur le modèle du cerveau humain et de sa plasticité qui vise à créer de nouvelles connexions synaptiques à partir d'expériences passées : c'est le sens de l'intelligence que de se reconfigurer en permanence en vue de l'action future. L'IA générative offre

¹ *Ibid.*, p. vi.

² *Ibid.*, p. xv.

³ *Ibid.*, p. 152.

ainsi l'expérience d'un apparaître permanent, une donation par esquisses qui n'est pas sans rappeler la description phénoménologique de l'être-au-monde. Pour Manovich, ce jaillissement constant de nouveauté permet un véritable renouvellement créatif, les modèles de génération donnant naissance à de nouveaux artefacts doté d'un statut propre :

En matière d'IA artistique, nous ne devons pas nous laisser aveugler par la manière dont ces systèmes sont entraînés. Certes, les modèles d'IA sont entraînés à partir d'œuvres d'art et d'artefacts culturels créés précédemment par l'homme. Cependant, leurs nouvelles productions ne sont pas des répliques mécaniques ou des simulations de ce qui a déjà été créé. À mon avis, il s'agit souvent d'artefacts culturels véritablement nouveaux, avec un contenu, une esthétique ou des styles inédits.¹

Mais justement, on retrouve là la véritable question de la créativité. Bien sûr, au sens propre, l'IA générative crée bien, on vient de le voir, des contenus absolument nouveaux. Cependant, à un niveau plus symbolique, la créativité désigne la capacité à s'éloigner des schémas classiques : produire, non pas de la nouveauté absolue à partir de rien, mais des perspectives nouvelles sur des choses préexistantes. Manovich n'est pas dupe : « Le simple fait d'être nouveau ne rend pas automatiquement quelque chose d'intéressant ou d'important sur le plan culturel ou social »². Si l'on veut faire preuve de charité vis-à-vis des critiques de l'IA générative, on peut considérer que le manque d'originalité dont elles lui font reproche ne portent pas une créativité de bas niveau, simple nouveauté, mais de haut niveau, intégrant la nécessité d'un intérêt significatif. Même si les images générées sont strictement nouvelles en ce qu'elles présentent un arrangement de *pixels* inédit, il est vrai que bon nombre de générations proposent des résultats très clichés, aseptisés, manquant sérieusement de caractère et de différence. De fait, le processus de génération est comme une partie de pêche dans l'océan immense des images possibles, et, à chaque fois qu'on lance sa ligne on est sûr de mordre. Il est donc aisé de produire à tour de bras du contenu. L'originalité authentique est en revanche bien plus difficile à atteindre. Puisque les modèles sont entraînés à modéliser une fonction d'approximation de la distribution d'un *dataset*, ils ont par défaut tendance à donner des résultats, certes différents, mais généralement très proches de ce *dataset*. Tout l'art de la génération consiste

¹ L. Manovich et E. Arielli, *Artificial Aesthetics*, op. cit., chap. 5 "From Representation to Prediction: Theorizing the AI Image", Traduction personnelle : « When it comes to artistic AI, we should not be blinded by how these systems are trained. Yes, AI models are trained on previously created human art and culture artifacts. However, their newly generated outputs are not mechanical replicas or simulations of what has already been created. In my opinion, these are frequently *genuinely new* cultural artifacts with *previously unseen content, aesthetics, or styles*. »

² *Ibid.*, Traduction personnelle : « simply being novel does not automatically make something culturally or socially interesting or significant »

alors à guider le modèle pour l'éloigner des lieux communs et l'entraîner sur des nouveaux chemins. En cela, l'IA générative ne s'oppose pas au travail véritable des artistes : tout est itératif, modifiable, « promptable », ouvrant un champ des possibilités inédit dans l'histoire de la figuration. Mais l'intelligence artificielle ne constitue pas pour autant une solution miracle au problème de la créativité : sans guide, sans effort sincère de la part de l'utilisateur, les modèles n'aboutiront qu'à un résultat médiocre et commun, pas plus digne d'intérêt que le travail d'un peintre ou d'un sculpteur paresseux qui, après avoir appris la structure des grandes œuvres, se contenterait de la répéter sans apporter de nouvelle interprétation. En cela, seul le prompteur qui s'engage dans une véritable relation avec le modèle, dans un véritable processus artistique peut être dit artiste et réaliser des œuvres dignes de ce nom.

Ce dialogue avec la machine peut être poussé jusqu'à faire d'elle une collaboratrice artistique. Suivant le modèle précurseur d'Harold Cohen qui travaillait de paire avec son programme *AARON*, les cinéastes qui entendent travailler avec l'IA générative doivent accepter une forme de partage créatif. Il faut, avec ses *prompts*, diriger le modèle comme on dirigerait un acteur en modifiant légèrement les indications après chaque prise d'après les observations de son jeu, de son comportement. Cela implique une bonne connaissance du modèle, savoir quelles sont ses tendances, quelle est sa palette de jeu, quelle sont les indications, les formulations qui permettent d'en tirer le meilleur résultat. C'est, on l'a vu, le principe même du *prompt engineering* que de développer un savoir des *prompts* — la manière de les structurer, les informations pertinentes, etc... — mais surtout un savoir du modèle qui est comme un interprète, mais pas n'importe lequel : un super interprète jouant non seulement le rôle des acteurs mais aussi celui du directeur de la photo, des machinistes, des costumiers et même du décors. En fin de compte, le modèle se retrouve à interpréter l'ensemble de ce qui se trouve à l'image, depuis le mouvement des personnages jusqu'au phénomènes physiques et biologiques. Dans cette collaboration, les *prompts* peuvent perdre une partie de leur caractère formel pour devenir un véritable échange, comme on échangerait avec un collaborateur humain. Les dialogues avec *ChatGPT* auquel on peut demander la génération de scénarios, de *prompts* pour d'autres modèles ou bien directement d'images prennent parfois des accents confondants d'humanité. Récemment, *Runway* a également introduit un *chat* qui rend l'interaction avec le modèle bien plus efficace : il est possible de revenir plus en détails sur les *prompts* pour les raffiner par le dialogue, poussant un cran plus loin le processus créatif. Si la génération de vidéos par IA semblait loin d'appartenir à la catégorie de la prise de vues de par son

absence d'espace et de temps réel, elle s'en rapproche au contraire dans cette dimension générative et collaborative. Les générations ne sont pas sans rappeler les « prises » du tournage. On fonctionne par itérations, par allers retours successifs. Comme le cinéaste qui refait encore et encore ses prises pour affiner le résultat au maximum, le prompteur génère sans cesse en modifiant ses *prompts* et ses *inputs* au fur et à mesure jusqu'à obtenir le résultat idéal. En ce sens l'IA permet au processus artistique de se dérouler pleinement, d'une manière parfois même plus riche que le tournage classique en prise de vues réelles en introduisant une méthode de réalisation circulaire. En effet, nous avons vu à quel point le scénario était remis au centre de la création des images — avec le *prompt*, le texte devient le sol dans lequel pousse directement les images, incitant les scénaristes à se faire déjà metteurs en scène en incluant dans leur traitement des indications de caméra, de lumières... Mais ce n'est pas seulement le scénario qui est remis au centre, ce sont en réalité toutes les étapes de réalisation d'un film qui sont condensées, concentrées autour de la génération. Si le *pipeline* de réalisation d'un film classique est envisageable comme une continuité linéaire avec différentes étapes successives dont les trois grandes catégories sont la pré-production, la production et la post-production, la réalisation d'un film par IA générative peut plutôt être envisagé de manière circulaire : la génération d'images au centre, et les différentes étapes, comme l'écriture ou le montage, s'agglutinant autour. Dans le cas du montage justement, celui-ci n'a plus vraiment de raison de venir après une étape de tournage, d'enregistrement, ou plutôt de génération, des *rushes*. Puisque le modèle est disponible à toute heure, présent sur la même machine que le logiciel de montage, les deux processus de création des *rushes* et de leur montage peuvent se dérouler en parallèle, se nourrissant l'un l'autre. Le cinéaste monte, petit à petit, les images dont il dispose et génère dans le même temps les plans exigés par le rythme et la continuité spatiale dans un processus itératif en temps réel. Ainsi, si l'analogie du tournage virtuel se heurte, pour le moment, à la non-persistance de l'espace généré par l'IA, voire même à la non-existence d'un quelconque espace, elle fonctionne en revanche particulièrement bien dans la dimension de nouveauté créatrice que permet le médium : à chaque itération, à chaque « prise », le cinéaste approfondit un peu plus la relation qui le lie à son interprète de façon à le pousser à la surprise et à la véritable originalité.

Finalement, la création de film par IA réduit l'équipe au minimum : dans son unité de base, elle ne comprend que le cinéaste dialoguant avec le modèle. Bien qu'une grande part du processus créatif d'un film repose sur la dimension collective et les échanges constants entre les dizaines voire centaines de personnes investies dans la production, la réduction de l'équipe à cette unité atomique

constitue une libération importante. Nombre des contraintes créatives du cinéma sont en réalité des contraintes techniques ou financières provenant de cette infrastructure énorme qu'impose la réalisation d'un film. L'animation 3D, si elle s'abstrait d'une certaine autorité du réel, impose toujours un processus de travail si long qu'il est difficile de réaliser un projet sans équipe. Désormais, avec l'IA générative c'est possible ; on peut se passer de cette infrastructure pour réaliser à l'envie des films seul avec son ordinateur, comme l'écrivain n'a besoin que de son stylo. Il est d'ailleurs notable que cette libération se fasse par le biais, en grand partie, de l'écriture, l'IA rapprochant en cela l'expérience du cinéaste de celle de l'écrivain. Bien sûr, la plupart des films réalisés avec des modèles génératifs continueront de l'être dans le cadre d'une importante structure humaine, et souvent pour le mieux, mais la possibilité pour le réalisateur de générer un film seul sera toujours là, garantissant sa liberté.

VII. Perspectives futures : cinéma de prise de vues virtuelles ou *artificial live action*

Générer des mondes...

Jusqu'à maintenant, nous avons examiné deux types principaux de synthèse d'image permettant un semblant de prise de vues. Le premier, incarné par les moteurs de rendu 3D temps réel comme *Unity* ou *Unreal Engine*, offre un haut degré d'immersion corporelle et de consistance spatiale : on peut s'y déplacer librement, interagir avec des objets, explorer des environnements persistants. En revanche, ce mode synthèse reste peu génératif : tout ce qui apparaît à l'écran doit être préalablement modélisé, codé ou paramétré. L'IA générative, à l'inverse, peut produire un flux ininterrompu de nouveautés visuelles ou narratives. Mais ces mondes ont un faible niveau d'immersion : ils manquent de stabilité et offrent une immersion spatiale très pauvre quand elle n'est pas inexistante. En un mot, ils ne permettent pas de perception dans un espace virtuel. Dans les deux cas, la « prise de vues virtuelles » — c'est-à-dire le fait d'habiter un point de vue situé dans un monde simulé — se heurte à une limite structurelle. Pour dépasser ce problème, il faut envisager des mondes virtuels dans un sens fort : combinant à la fois la consistance tridimensionnelle d'un moteur 3D et la générativité ouverte des modèles d'IA. Nous avons vu dans le chapitre 2 que David Chalmers définissait la réalité virtuelle comme un espace « immersif, interactif, et généré par ordinateur »¹. De cette définition, Chalmers fait découler celle de « monde virtuel », qui repose également sur ces trois critères mais exige un degré moindre d'immersivité : il suffit d'une spatialité convaincante — par exemple, celle fournie par un moteur 3D sur écran. En revanche, l'interactivité est cruciale : c'est elle qui distingue un monde virtuel de toute autre simulation informatique. Chalmers résume ainsi cette distinction :

Tout monde virtuel implique un espace, que nous pouvons en principe interpréter comme une simulation d'un espace physique hypothétique. Dans ce sens large, tout monde virtuel implique une simulation informatique. Qu'en est-il de l'inverse ? À strictement parler, toutes les simulations informatiques ne sont pas des mondes virtuels. Il existe des simulations non interactives, comme les

¹ D. Chalmers, *Reality+*, *op. cit.*, p. 193-194.

simulations standard de la formation des galaxies, qui n'interagissent pas du tout avec les utilisateurs. Puisqu'elles ne sont pas interactives, elles ne répondent pas à la définition des mondes virtuels.¹

Un monde virtuel doit donc posséder une forme d'autonomie lui permettant de réagir aux actions de l'utilisateur. On peut distinguer trois niveaux de simulation d'un monde virtuel selon la profondeur de cette autonomie.

Le premier niveau désigne les simulations déterministes à l'image des scènes 3D avec lesquelles interagit directement un infographiste dans un studio d'animation ou d'effets spéciaux. Ici, l'autonomie est plutôt une forme d'automatisme : l'ensemble des phénomènes et des mouvements sont contrôlés directement par l'utilisateur sans que le programme ne dispose d'une marge de manœuvre importante, celle-ci se réduisant le plus souvent à une forme d'interpolation. C'est le cas de la plupart des moteurs de rendu 3D qui simulent des lois physiques comme le tracé de la lumière ou le mouvement de corps en suivant des règles déterministes. Le programme se contente de suivre des directives sans véritablement participer au processus de création. C'est un monde d'une intensité très faible, permettant un niveau d'interactivité utilitaire : la simulation est avant tout un outil. Comme nous l'avons déjà évoqué, la création d'images de synthèse avec des moteurs 3D est une création automatisée dans laquelle le logiciel joue un rôle d'intermédiaire. Dès lors, plus l'automatisme est poussé, plus la création est facilitée. Les modes de travail procéduraux permettent, par exemple, de créer de nouvelles combinaisons à partir d'éléments déjà existants pour donner lieu à de nouvelles entités. Le logiciel *Houdini* est connu pour son système nodal qui permet un haut niveau de création procédurale, mais d'autres logiciels développent ce genre d'architecture à l'image de *Blender* et de ses *geometry nodes*. En assemblant entre eux des blocs qui effectuent chacun des opérations particulières, il est possible d'automatiser une grande part du travail de modélisation et de créer de multiples déclinaisons d'un modèle 3D, par exemple un paysage, en faisant varier l'agencement de ses composantes. Quoi qu'il en soit, rien ne bouge « tout seul » dans ce genre de simulation sans intervention directe de l'utilisateur : chaque position, chaque objet, chaque lumière est entièrement sous le contrôle de l'utilisateur. De telles simulations sont plus des espaces que des mondes au sens plein.

¹ *Ibid.*, p. 54., Traduction personnelle : « Any virtual world involves a space, which we can in principle interpret as a simulation of a hypothetical physical space. In this broad sense, any virtual world involves a computer simulation. What about the converse? Strictly speaking, not all computer simulations are virtual worlds. There are noninteractive simulations, like standard simulations of galaxy formation, that do not interact with users at all. Because they aren't interactive, they don't meet the definition of virtual worlds. »

Le second niveau est celui des simulations interactives dont les jeux vidéo sont les meilleurs représentants. Ici, l'agencement du monde virtuel et les mouvements qui le parcourent, s'ils peuvent être causés par l'utilisateur, lui échappent en grande partie. Des phénomènes comme le balancement d'arbres virtuels, l'écoulement d'un cours d'eau ou encore les comportements des *Non-Player Character* se produisent de manière indépendante, procurant à l'utilisateur qui n'a pas la main sur leur déroulement l'impression d'être *jeté* dans un monde virtuel. Mais tous ces mouvements sont préexistants, pré-fabriqués comme d'ailleurs l'ensemble de la simulation. Les *workflows* procéduraux, que nous venons de mentionner, sont souvent utilisés pour donner plus de souplesse à ces simulations qui autrement apparaîtraient purement déterministes. La qualité du niveau de simulation d'un jeu dépend alors de la complexité du système qu'il déploie. Certains jeux comme *No Man's Sky* permettent d'explorer une infinité de planètes, générées en permanence à partir de composantes pré-existantes. D'autres comme *Red Dead Redemption 2* sont notoirement connus pour offrir une pléthore de scénarios expérimentables et d'actions possibles, donnant la sensation d'un monde ouvert. Les chiffres parlent d'eux mêmes : le scénario du jeu fait près de 2000 pages, contient 450 personnages, 500 000 lignes de dialogues, 80 pages de script par *NPC*, 2200 jours de tournage en *motion capture* pour 300 000 animations jouées par près de 1000 acteurs... Et tous ces éléments se multiplient entre eux, de manière souvent aléatoire, pour donner une quantité astronomique de cas de figure différents. Résultat : 65 heures de jeu pour le scénario principal. Dans *Reality +*, Chalmers classe ces différents objets virtuels peuplant les jeux vidéo selon cinq catégories en fonction de leur degré d'agentivité¹. D'abord les « objets décoratifs » : objets virtuels qui ont le moins de pouvoir causal, ils servent seulement à « décorer » l'espace ; ensuite les « objets solides » qui ne peuvent pas être pénétrés par d'autres objets ; puis les « objets mobiles » qui peuvent bouger ou changer d'aspect (en principes ces objets reposent sur des simulations physiques) ; suivis des « objets spéciaux », dotés de pouvoirs qui, eux, échappent au moteur physique (par exemple, pour un *NPC* être doué d'une pseudo intériorité...) ; et enfin, les « objets animés » qui sont contrôlés par l'utilisateur, en particulier son avatar. Plus l'agentivité des objets qui peuplent une telle simulation augmente, plus celle-ci s'enrichit et se rapproche d'un véritable monde virtuel. Mais il n'en reste pas moins que cette simulation est entièrement dépendante d'un programme : toute la procéduralité et l'aléatoire, aussi riches et complexes soient-ils, restent bornés par les limites du code initial.

¹ *Ibid.*, p. 200.

Le troisième et dernier niveau d'autonomie est celui des simulations génératives dans lesquelles, justement, les éléments du monde ne sont pas précodés mais créés en temps réel par un modèle génératif. En effet, à la différence du rendu en temps réel qui calcule des images d'après une réalité préexistante, la génération en temps réel fait entrer ses objets dans l'existence au moment même où ils deviennent visibles. Dans une telle simulation, chaque objet, chaque décor surgit à tout instant d'un néant informatique. Ce procédé résout le manque de flexibilité des deux niveaux précédents, mais pose le problème de la persistance : comment garantir que ce qui a été généré puisse réapparaître ultérieurement de manière reconnaissable ? Les premier et second niveaux de simulation existent avec les moteurs de 3D *offline* et *realtime*, mais qu'en est-il du troisième niveau d'autonomie, le seul qui puissent permettre d'accéder à la simulation de mondes virtuels au sens plein ? Si les défis qu'il pose semblent correspondre aux potentialités de l'IA générative, nous avons vu que les modèles de diffusion de vidéo, limités par l'absence de spatialité et de persistance, n'y suffisent pas. Un nouveau type de modèle commence à émerger qui semble répondre à ces problématiques : les *world models*.

L'IA, à travers la poursuite de l'*AGI* (*Artificial General Intelligence*), cherche une autonomie de plus en plus grande et surtout de plus en plus complète, c'est-à-dire capable d'opérer dans l'environnement le plus large regroupant tous les cas de figure, tous les possibles : le monde lui-même. Pour ce faire, les modèles doivent parvenir à une compréhension globale. Nous avons vu que les algorithmes d'IA générative réalisaient des fonctions d'approximation : ils font en quelque sorte la cartographie de l'espace des possibles. Pour les *LLMs*, cet espace est celui du langage. Pour les modèles de génération d'images et de vidéos, les images s'ajoutent au langage : ces modèles constituent en cela une forme plus avancée de cartographie du monde que les modèles de langage. Et cela va de plus en plus loin. Dans leur poursuite de la persistance spatio-temporelle, les modèles de vidéos comme *Aleph* de *Runway* ou *Sora* d'*Open AI* en viennent à intégrer des règles géométriques et physiques : ils développent une représentation interne de la structure du monde. *Sora* parvient, par exemple, à maintenir la persistance d'un objet au long d'une vidéo pendant laquelle il est occulté — un objet visible au début de la séquence, puis dissimulé par un second élément passant au premier plan, réapparaît à la fin de la séquence lorsque l'objet du premier plan s'en est allé — ou encore à représenter des interactions cohérentes au sein du monde — un personnage qui marche dans la neige et laisse des traces de pas, un homme qui croque dans un burger et laisse une marque de dents, etc... Ces améliorations constantes vont dans le sens d'une connaissance de plus en plus globale du monde : à mesure que la représentation interne du modèle

se structure, les éléments générés acquièrent une forme de causalité entre eux. Mais malgré cette compréhension sans cesse améliorée de la physique et de la géométrie, les modèles de génération de vidéo peinent encore à produire une connaissance vraiment unifiée et structurée du monde. On a vu dans le chapitre précédent la pauvreté de l'expérience immersive fournie par les modèles actuels et leur manque de consistance. Surtout, ces modèles ne permettent pas une interaction en temps réel comme un moteur de jeu vidéo : on ne peut pas se déplacer dans un espace et interagir avec des éléments de cet espace. C'est que la génération de mondes interactifs ne se résume pas seulement au problème de la génération de vidéo en temps réel :

À première vue, la simulation des mondes interactifs des jeux vidéo peut sembler similaire à la génération vidéo. Cependant, la simulation de mondes interactifs est bien plus qu'une simple génération vidéo très rapide. La nécessité de conditionner la génération à un flux d'actions en entrée qui n'est disponible que pendant la génération elle-même remet en cause certains postulats des architectures de modèles de diffusion existants.¹

Face à cette difficulté, les *world models* s'imposent comme la solution de l'avenir, la prochaine étape de l'IA générative dans sa modélisation du monde, après les *LLMs* et les modèles de génération d'images et de vidéos. Un *world model* est un modèle de *deep learning* qui représente dynamiquement un monde virtuel en générant sa géométrie, sa physique et des interactions possibles. Voici la définition qu'en fournit le constructeur de cartes graphiques *Nvidia* :

Les « World Models » sont des modèles d'IA générative qui comprennent la mécanique du monde réel, y compris la physique et les propriétés spatiales. Ils utilisent des données d'entrée, notamment du texte, des images, des vidéos et des mouvements, pour générer des vidéos. Ils comprennent les propriétés physiques des environnements du monde réel en apprenant à représenter et à prédire des mécanismes tels que le mouvement, la force et les relations spatiales à partir de données sensorielles.²

Les *LLMs* génèrent du texte, les modèles de diffusion génèrent des images ou des vidéos et les *world models* génèrent des mondes. Leurs générations se présentent toujours sous la forme d'un flux vidéo mais il ne faut pas s'arrêter là : sous la surface visuelle, ces modèles calculent et prennent

¹ VALEVSKI Dani, LEVIATHAN Yaniv, ARAR Moab, et al., « Diffusion models are real-time game engines », in *arXiv preprint arXiv:2408.14837*, 2024., Traduction personnelle : « At a glance, simulating the interactive worlds of video games may seem similar to video generation. However, interactive world simulation is more than just very fast video generation. The requirement to condition on a stream of input actions that is only available throughout the generation breaks some assumptions of existing diffusion model architectures. »

² Texte original sur le site de NVIDIA : <https://www.nvidia.com/en-us/glossary/world-models/>

en compte l'architecture sous-jacente du monde virtuelle. Ce mode de fonctionnement ouvre des possibilités nouvelles, en particulier dans les champs qui nous intéressent. D'abord, les *world models* parviennent à une représentation tridimensionnelle et stable des mondes virtuels dans laquelle les utilisateurs peuvent se déplacer avec les touches du clavier ou la souris, ouvrant la voie à un contrôle de caméra et à une expérience perceptive proche de celle des moteurs 3D. Par ailleurs, un *world model* complet peut assurer la cohérence narrative en suivant l'état du monde et des personnages, facilitant l'élaboration de narrations procédurales ou interactives. Un tel modèle de monde maintient en mémoire la situation (personnages, objets clés, événements passés) pour générer des scénarios cohérents. Ici, plus qu'un espace 3D, nous avons un monde capable de réagir et même d'agir, de produire une activité et une narration. Enfin, les *world models* permettent de créer automatiquement des environnements 3D (terrains, architectures, végétation...) à la demande pour des jeux vidéo ou toute autre expérience interactive en respectant un style ou des contraintes.

Parmi les *world models* existants, on peut distinguer deux grandes approches : *video-based* ou *3D-based*. La première approche génère une vidéo, un flux continu d'images pixelisées, permettant une très grande flexibilité. La seconde approche génère une véritable géométrie 3D (*vertices*, côtés, surfaces) plus appropriée à une bonne consistance spatiale. Parmi les *world models* basés sur de la vidéo on peut distinguer *GameNGen*, *Oasis* de *Decart Labs*, *Genie 2* puis *Genie 3* de *Google*, ou encore *Muse* de *Microsoft*. *GameNGen*, présenté en 2024 dans l'article *Diffusion models are real-time game engines*¹, constitue l'un des premiers *video world models* à permettre une interaction en temps réel. Entraîné avec des séquences de *gameplay* du très classique jeu *Doom*, le réseau de neurones de *GameNGen* joue le rôle d'un moteur de rendu 3D en temps réel offrant une expérience de jeu à 20 *fps* quasiment indistinguishable du jeu original. Dans la lignée des modèles entraînés sur des jeux vidéo, *Oasis* par *Decart Labs* constitue un autre exemple notable. Le modèle permet de jouer à une version générée par IA de *Minecraft* directement dans un navigateur internet en reproduisant des actions classiques comme couper du bois, construire ou combattre des zombies. Mais c'est véritablement avec les modèles de *Google DeepMind Genie 2* en 2024, puis *Genie 3* présenté en août 2025, que les *world models* prennent une nouvelle dimension en permettant aux utilisateurs de générer leur propre environnement 3D interactif à partir d'une image ou de *prompts*. Si *Genie 2* était encore limité à la génération de séquences d'une dizaine de secondes, *Genie 3* offre des résultats impressionnants de persistance spatio-temporelle dans des générations de plusieurs minutes. Et surtout, cette dernière mise-à-jour du modèle fonctionne en temps réel : il est désormais

¹ D. Valevski, Y. Leviathan, M. Arar, et al., « Diffusion models are real-time game engines », art cit.

possible d'interagir avec des environnements tridimensionnels à l'esthétique et au *design* inédits. Enfin, parmi les avancées notables, on pourrait citer le modèle *WHAM* introduit par *Microsoft* début 2025 qui permet de relever trois défis de taille dans la génération d'univers interactifs : la cohérence, la diversité et la persistance :

a, Cohérence : une séquence générée doit être cohérente dans le temps et avec les mécanismes du jeu. [...] b, Diversité : le modèle doit produire des séquences nombreuses et variées qui reflètent différents résultats potentiels afin de favoriser la pensée divergente. [...] c, Persistance : le modèle doit conserver les modifications apportées par l'utilisateur aux visuels du jeu et aux actions du contrôleur, en les assimilant dans la séquence de jeu générée.¹

Microsoft a même rendu accessible un « Wham demonstrator » offrant la possibilité aux développeurs de tester le modèle avec différents *inputs* comme des images.

Du côté des *world models* basés sur la 3D, on peut mentionner *World Labs*, *Intangible* ou encore *Third Dimension*. La start-up *World Labs* fondée par Fei-Fei Li, l'instigatrice de la base de données *ImageNet* qui avait permis une avancée majeure dans la reconnaissance d'images, cherche à concevoir des « *LWM* » — *Large Worlds Models* — capable de « percevoir, générer et interagir avec des mondes en 3D »². Pour Li, l'intelligence spatiale est la forme d'intelligence primordiale, première dans notre manière d'interagir avec le monde, rejoignant en cela la conception de la primauté du corps chez Merleau-Ponty. Selon la chercheuse, l'IA qui s'est jusqu'à maintenant principalement concentrée sur le langage doit s'attaquer à ce type d'intelligence plus complexe pour progresser vers un nouveau pallier. Li fait donc le pari de la 3D pour atteindre l'intelligence spatiale:

Pour aller au-delà des capacités des modèles actuels, nous avons besoin d'une IA spatialement intelligente, capable de modéliser le monde et de raisonner sur les objets, les lieux et les interactions dans un espace et un temps en 3D. [...] Notre objectif est de faire passer les modèles d'IA du plan 2D des pixels à des mondes entièrement en 3D, tant virtuels que réels, en leur conférant une intelligence spatiale aussi riche que la nôtre.³

¹ KANERVISTO Anssi, BIGNELL Dave, WEN Linda Yilin, et al., « World and Human Action Models towards gameplay ideation », in *Nature*, n° 8051, vol. 638, 1 février 2025, p. 656-663., Traduction personnelle : « **a**, Consistency: a generated sequence should be consistent over time and with game mechanics. [...] **b**, Diversity: the model should produce numerous, diverse sequences that reflect different potential outcomes to support divergent thinking. [...] **c**, Persistency: the model should persist user modifications to the game visuals and controller actions, assimilating them into the generated gameplay sequence. »

² Voir le site internet de *World Labs* : <https://www.worldlabs.ai/about>

³ *Ibid.*

World Labs a déjà présenté des résultats probants de génération, à partir d'une simple image, d'une scène 3D dans lequel on peut se déplacer comme dans l'interface d'un moteur de rendu.

Quelle que soit la méthode employée, basée sur de la vidéo ou sur de la géométrie 3D, les *world models* constituent une nouvelle étape dans l'IA générative, un nouveau niveau de modélisation du monde : non plus seulement à l'échelle du langage ou des images, mais de la physique et de la matérialité. Les chercheurs derrière *GameNGen* promettent un nouveau paradigme créatif :

Aujourd'hui, les jeux vidéo sont programmés par des humains. GameNGen est une preuve de concept pour une partie d'un nouveau paradigme dans lequel les jeux seraient des poids d'un modèle neuronal, et non des lignes de code. GameNGen montre qu'il existe une architecture et des poids de modèle permettant à un modèle neuronal d'exécuter efficacement un jeu complexe (DOOM) de manière interactive sur du matériel existant. Bien que de nombreuses questions importantes restent en suspens, nous espérons que ce paradigme pourra apporter des avantages importants. Par exemple, le processus de développement des jeux vidéo dans le cadre de ce nouveau paradigme pourrait être moins coûteux et plus accessible, les jeux pouvant être développés et modifiés à l'aide de descriptions textuelles ou d'images de référence.¹

Pour l'instant, ces modèles se heurtent encore à de nombreuses limites : la cohérence et la persistance — physique, spatiale ou narrative — demeurent des problèmes importants, surtout pour de longues générations, les modèles disposant d'une mémoire limitée ; la performance est un autre facteur bloquant, les meilleurs modèles atteignant 20 *fps*, très loin des moteurs de jeu comme *Unreal Engine*, et ce au prix d'une piètre qualité d'image et d'une puissance de calcul importante ; les données d'entraînement manquent pour modéliser des mondes complets, surtout dans le cas des modèles basés sur la 3D ; enfin, les *world models*, tout comme les modèles de génération de vidéos ont un fonctionnement très opaque et une faible granularité, offrant peu de contrôle pour *designer* avec précision ces mondes virtuels. Mais malgré l'ampleur de ces difficultés, il ne fait pas vraiment de doute qu'elles seront surmontées, petit à petit, jusqu'à permettre de générer des jeux entiers, à l'image des 65 heures de *Red Dead Redemption II*, voire des simulations infinies. En effet, là où un système procédural pouvait procurer au joueur l'impression d'une grande liberté, voire d'un monde

¹ D. Valevski, Y. Leviathan, M. Arar, et al., « Diffusion models are real-time game engines », art cit., Traduction personnelle : « Today, video games are programmed by humans. GameNGen is a proof-of-concept for one part of a new paradigm where games are weights of a neural model, not lines of code. GameNGen shows that an architecture and model weights exist such that a neural model can effectively run a complex game (DOOM) interactively on existing hardware. While many important questions remain, we are hopeful that this paradigm could have important benefits. For example, the development process for video games under this new paradigm might be less costly and more accessible, whereby games could be developed and edited via textual descriptions or examples images. »

ouvert, au prix d'une quantité astronomique d'éléments préfabriqués et reconfigurés en permanence pour donner un semblant de nouveauté, les limites repoussées finissaient toujours par ressurgir à un moment ou à un autre, dans l'impossibilité d'ouvrir une porte ou le caractère répétitif d'une ligne de dialogue, violent rappel de l'illusion. Les *world models* sont une promesse de régler la finitude des jeux vidéo. À partir d'une même image de départ, *Genie 3* génère de la diversité, des interactions différentes à chaque itération, permettant une expérience véritablement unique pour chaque joueur et pour chaque partie. Surtout, ces modèles rendent disponibles à l'interaction l'ensemble des objets virtuels : finis les portes qui ne s'ouvrent pas ou les *NPC* en boucle sur la même phrase. Les *world models* promettent de véritables mondes virtuels qui ne sont plus de simples décors. Plus largement qu'aux jeux vidéo, ils s'attaquent au mythe de la simulation : recréer un monde au sens plein de manière artificielle. Mais pour être vraiment des mondes, encore faut-il que ces espaces virtuels cochent les quatre critères phénoménologiques. Ce sont déjà des totalités autonomes, inachevées et horizon de la perception de celui qui s'y plonge. Mais il leur reste à être des lieux d'intersubjectivité : il faut qu'ils soient peuplés.

...Peuplés d'agents

L'idée qu'une intelligence artificielle puisse constituer un sujet est présente dès les origines de la discipline mais elle prend une nouvelle actualité avec l'émergence du paradigme de l'agent. En IA, « un agent intelligent est une entité qui perçoit son environnement, agit de manière autonome pour atteindre des objectifs et peut améliorer ses performances grâce à l'apprentissage automatique ou à l'acquisition de connaissances. »¹. D'une manière générale, un agent désigne toute entité capable de percevoir son environnement et d'agir en fonction. Dans ce sens les humains comme les animaux sont des agents, sans cesse agités de cet aller-retour entre perception et action qui est à la source de l'agentivité. Dans son livre qui retrace l'histoire de l'IA, Michael Wooldridge décrit l'émergence de ce paradigme de l'agent qui apparaît dans les années 1980 comme une solution aux difficultés rencontrées jusqu'à là par la recherche :

Un agent était conçu comme un système d'IA « complet » dans le sens où il s'agissait d'une entité autonome et indépendante, située dans un environnement donné et exécutant une tâche spécifique

¹ Wikipedia contributors, "Intelligent agent," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=Intelligent_agent&oldid=1304262187 (accessed August 15, 2025).

pour le compte d'un utilisateur. [...] Un consensus s'est imposé sur le fait que les agents devaient posséder trois capacités importantes. Premièrement, ils devaient être réactifs : ils devaient être en phase avec leur environnement et capables d'adapter leur comportement de manière appropriée lorsque des changements dans l'environnement l'exigeaient. Deuxièmement, ils devaient être proactifs : capables de travailler systématiquement pour accomplir la tâche qui leur était confiée pour le compte de leur utilisateur. Enfin, les agents étaient sociaux : capables de travailler avec d'autres agents lorsque cela était nécessaire.¹

Le concept d'agent intelligent est séduisante et finit par l'emporter auprès des chercheurs comme du grand public, l'agent représentant le modèle paradigmatique de l'IA, le but à atteindre. En effet, lorsqu'on imagine une intelligence artificielle, on pense d'abord à un assistant virtuel capable de nous seconder dans notre travail comme dans notre vie. La pop culture regorge d'exemples de ce genre, depuis l'ordinateur Hal à bord de la station spatiale de 2001, l'odyssée de l'espace, à la compagne virtuelle de Joaquin Phoenix dans Her. Mais dans les années 80, l'agent intelligent est encore hors de portée. Ce qui définit un agent, c'est sa capacité à percevoir dans un environnement qui peut être le monde lui-même, et fournir des réponses appropriées. Dès lors, l'agent doit être en mesure de percevoir et de réagir convenablement à des situations potentiellement toujours nouvelles et imprévisibles. La capacité de mener à bien des tâches pour lesquelles des IA n'ont pas été programmées devient une réalité à partir des années 2000 et le tournant du machine learning, plus précisément du deep learning. Il devient alors possible d'envisager des agents intelligents doués d'une autonomie de plus en plus importante. Dans cette course vers l'agent idéal, de nombreuses échelles de mesure de l'autonomie ont été établies pour servir de guide à la recherche. Dans leur article, Levels of Autonomy for AI Agents² paru en 2025, Kevin Feng et ses collègues proposent cinq niveaux d'autonomie en fonction du rôle joué par l'utilisateur. Cela fait sens de se baser sur ce critère puisque l'autonomie se définit par la capacité d'un agent à agir sans supervision d'un opérateur humain. Ainsi, le niveau 1 est celui de « l'utilisateur comme opérateur » : « L'utilisateur dirige et prend les décisions, l'agent agit ». Le niveau 2 celui « l'utilisateur comme collaborateur » :

¹ M.J. Wooldridge, *A brief history of artificial intelligence*, op. cit., p. 82., Traduction personnelle : « An agent was intended to be an AI system that was “complete” in the sense that it was a self-contained, autonomous entity, situated in some environment and carrying out some specific task on behalf of a user. [...] a consensus emerged that agents had to have three important capabilities. First, they had to be reactive: they had to be attuned to their environment and able to adapt their behavior appropriately when changes in the environment demanded it. Second, they had to be proactive: capable of systematically working to achieve their given task on behalf of their user. Finally, agents were social: capable of working with other agents when required. »

² FENG K. J., McDONALD David W. et ZHANG Amy X., « Levels of Autonomy for AI Agents », in *arXiv preprint arXiv:2506.12469*, 2025.

« L'utilisateur et l'agent planifient, délèguent et exécutent leurs tâches de manière collaborative ». Le niveau 3 de « l'utilisateur comme consultant » : « L'agent prend l'initiative, mais consulte l'utilisateur pour connaître son expertise et ses préférences. ». Le niveau 4 de « l'utilisateur comme approbateur » : « L'agent n'interagit avec l'utilisateur que dans des scénarios risqués ou prédéfinis ». Enfin, le niveau 5 est celui de « l'utilisateur comme observateur » : « L'agent fonctionne en toute autonomie sous la surveillance de l'utilisateur ». À mesure que l'agent gagne en autonomie, il est amené à améliorer la profondeur de sa perception et de son action. Si les deux premiers niveaux impliquent un environnement de perception relativement simple et des interactions de premier ordre, i.e. des interactions directes sans réflexivité, les niveaux suivants mettent en œuvre un environnement de plus en plus complexe, non-déterministe, jusqu'à se confondre avec le monde, et des interactions de haut-niveaux, impliquant différents degrés de réflexivité et une planification importante pour développer des stratégies d'action sophistiquées. Cette complexification croissante va de pair avec le développement de world models, des modélisations du monde par l'agent. Ici, le terme de « world models » diffère légèrement de celui que nous avons employé dans la partie précédente pour désigner les modèles de génération d'environnements virtuels. Mais au fond le principe est le même : il s'agit pour le modèle de génération d'environnements comme pour l'agent intelligent de développer une structure interne et consistante du monde. L'article *Recurrent World Models Facilitate Policy Evolution*¹ de David Ha et Jürgen Schmidhuber paru en 2018, l'un des premiers à mentionner les *worlds models*, propose d'ailleurs une méthode d'entraînement qui associe la génération d'environnements virtuels et l'entraînement d'un agent intelligent capable d'évoluer dans ce monde. Les agents « apprennent dans leurs propres rêves », ils génèrent eux-mêmes le monde dans lequel ils perçoivent et agissent et, de ce fait, structurent leurs propres « world models ». Les auteurs s'inspirent en cela directement de la manière dont nous, humains, construisons une représentation mentale du monde pour mettre en forme notre perception et notre action :

Les êtres humains développent un modèle mental du monde basé sur ce qu'ils sont capables de percevoir avec leurs sens limités. Les décisions et les actions que nous prenons sont basées sur ce modèle interne. [...] Pour traiter la grande quantité d'informations qui circulent dans notre vie quotidienne, notre cerveau conçoit une représentation abstraite des aspects spatiaux et temporels de ces informations. Nous sommes capables d'observer une scène et d'en retenir une description

¹ HA David et SCHMIDHUBER Jürgen, « Recurrent World Models Facilitate Policy Evolution », s.l., Curran Associates, Inc., 2018, vol. 31/.

abstraite. Des éléments suggèrent également que ce que nous percevons à chaque instant est régi par les prédictions de notre cerveau sur l'avenir, en fonction de notre modèle interne. Une façon de comprendre le modèle prédictif de notre cerveau serait de considérer qu'il ne s'agit pas seulement de prédire l'avenir en général, mais plutôt de prédire les données sensorielles futures en fonction de nos actions motrices actuelles. Nous sommes capables d'agir instinctivement sur la base de ce modèle prédictif et d'adopter des comportements réflexifs rapides lorsque nous sommes confrontés à un danger, sans avoir besoin de planifier consciemment une ligne de conduite.¹

Selon cette théorie, le modèle mental que nous formons du monde nous permet d'agir par réflexe, sans avoir à traiter l'information de manière abstraite et objective. Cela n'est pas sans rappeler le modèle décrit par Merleau-Ponty : le corps propre est d'abord dans une spatialité de situation, tourné vers l'action. Tout l'enjeu de l'entraînement des agents artificiels consiste alors à leur permettre de développer un modèle du monde dans lequel ils évoluent. C'est par l'apprentissage d'une telle structure, des dynamiques et des relations causales qui traversent le monde, que les agents seront en mesure de percevoir, d'agir et de s'adapter efficacement.

Cette intuition est également partagée par Yann LeCun, l'un des pères fondateurs du *deep learning* et inventeur des réseaux de neurones. Dans son article paru en 2022, *A Path Towards Autonomous Machine Intelligence*², autant d'ordre philosophique que scientifique en ce qu'il possède une dimension programmatique importante, LeCun expose ses intuitions et sa vision pour une future direction de la recherche en intelligence artificielle. Pour le chercheur, la principale différence entre les IA et les êtres vivants comme les humains ou les animaux réside justement dans la capacité de ces derniers à former des représentations du monde. Là où les modèles nécessitent une quantité astronomique d'information pour fonctionner correctement — un programme de voiture autonome doit avoir été entraîné sur tous les cas possibles en grande quantité, au point que le modèle aura déjà rencontré, à de nombreuses reprises, même les cas les plus rares — les humains et les animaux « semblent capables d'acquérir d'énormes quantités de connaissances générales sur le fonctionnement du monde grâce à l'observation et à un nombre incroyablement faible d'interactions,

¹ *Ibid.*, Traduction personnelle : « Humans develop a mental model of the world based on what they are able to perceive with their limited senses. The decisions and actions we make are based on this internal model. [...] To handle the vast amount of information that flows through our daily lives, our brain learns an abstract representation of both spatial and temporal aspects of this information. We are able to observe a scene and remember an abstract description thereof. Evidence also suggests that what we perceive at any given moment is governed by our brain's prediction of the future based on our internal model. One way of understanding the predictive model inside of our brains is that it might not be about just predicting the future in general, but predicting future sensory data given our current motor actions. We are able to instinctively act on this predictive model and perform fast reflexive behaviours when we face danger, without the need to consciously plan out a course of action. »

² LeCUN Yann et COURANT, « A Path Towards Autonomous Machine Intelligence Version 0.9.2, 2022-06-27 », s.l., s.n., 2022.

de manière indépendante et sans supervision ».¹ Entraîner des IA pour qu'elles acquièrent des représentations du monde constitue ainsi le gros défi à venir : « On peut affirmer que la conception d'architectures et de paradigmes d'apprentissage pour le modèle mondial constitue le principal obstacle à la réalisation de progrès réels dans le domaine de l'IA au cours des prochaines décennies »². Et le défi est d'autant plus important que la modélisation du monde « externe » — i.e. la structure directement accessible dans l'observation et qui entre en jeu dans les interactions de premier ordre — ne suffit pas. Pour permettre une interaction efficace avec d'autres agents (humains, animaux ou artificiels), les agents intelligents doivent également être en mesure de concevoir les représentations de leurs interlocuteurs. Il leur faut donc développer une *Theory of Mind (ToM)*, condition pour des interactions de haut niveau impliquant de la réflexivité — agir en fonction de croyances à propos de croyances, elles-mêmes à propos de croyances, etc... Le défi est donc de taille. Il se résume à trois grandes questions posées par LeCun dans son article : 1) Comment les machines peuvent-elles apprendre de manière aussi efficace que les animaux et les humains ?, 2) Comment les machines peuvent-elles apprendre à raisonner et à planifier ?, et enfin, 3) Comment les machines peuvent-elles concevoir des représentations et des plans d'action à plusieurs niveaux d'abstraction ?

Ces questions ne sont pas nouvelles et, parmi les tentatives de réponses qui y ont été apportées tout au long de l'histoire de la recherche, on peut distinguer deux grandes catégories : d'une part, l'IA symbolique, concentrée sur la manipulation de symboles, et, d'autre part, l'IA incarnée ou incorporée — « embodied AI » en anglais — tournée vers des modèles dotés d'un corps, comme des robots, et de ce fait capable de percevoir et d'agir dans un environnement physique. Historiquement, l'IA symbolique se réfère à une architecture logique permettant aux modèles de manipuler des symboles selon des règles explicites. Avec le *machine learning*, cette définition change sensiblement : si les modèles n'obéissent plus à des règles explicites, ils reposent encore, en revanche, sur la manipulation de symboles (images, mots, tokens...) qui ne sont pas ancrés dans une expérience vécue du monde. Dans son acception large, le paradigme symbolique d'une intelligence désincarnée l'a emporté jusqu'à maintenant, constituant une majeure partie de la recherche en intelligence artificielle. C'est particulièrement flagrant en ce moment avec les modèles

¹ *Ibid.*, Traduction personnelle : « seem able to learn enormous amounts of background knowledge about how the world works through observation and through an incomprehensibly small amount of interactions in a task-independent, unsupervised way. »

² *Ibid.*, Traduction personnelle : « Arguably, designing architectures and training paradigms for the world model constitute the main obstacles towards real progress in AI over the next decades. »

génératifs : les systèmes de traitement du langage naturel comme les *LLMs* s'appuient aujourd'hui sur une conception purement symbolique du langage. Et, en effet, les modèles désincarnés se sont révélés une intuition fructueuse à suivre en comparaison avec les difficultés rencontrées par les IA incarnées. Mais bien que les architectures symboliques, et les *LLMs* en particulier, offrent des résultats exceptionnels dépassant toutes les attentes, elles semblent encore loin d'une véritable compréhension. Sur ce point, elles se heurtent à un obstacle majeur : le « problème de l'ancrage des symboles ». Ce « grounding problem » est un problème fondamental de la sémantique « qui concerne la circularité dans la définition de la signification d'un symbole par le biais d'autres symboles »¹. Le problème est particulièrement vif dans le langage : si celui-ci est un ensemble de symboles et de combinaisons de symboles suivants des règles de syntaxe, alors d'où viennent ces symboles en premier lieu ? Dans le cas des systèmes de traitement du langage naturel qui reposent sur une conception symbolique, le problème de l'ancrage est purement et simplement ignoré : les symboles sont ancrés dans d'autres symboles qui sont eux-mêmes ancrés dans d'autres symboles, et ainsi de suite dans une régression infinie. De nombreuses voix critiques se basent sur ce phénomène pour défendre une absence de compréhension véritable de la part des systèmes symboliques qui se contentent de manipuler de l'information sans en saisir le sens profond. Le problème ressort clairement dans l'expérience de pensée de la « chambre chinoise » développée par le philosophe américain John Searle. Celui-ci imagine une chambre dans laquelle est enfermé un individu dont la seule interaction avec le monde consiste à recevoir des messages en chinois et à y fournir des réponses, toujours en chinois. Cet individu ne parle pas chinois, mais il dispose d'un manuel de règles syntaxiques, si complet et si précis, qu'il lui permet de répondre correctement à chacun des messages qu'il reçoit. Au point qu'il passe pour un véritable sinophone aux yeux de ses interlocuteurs extérieurs. Pourtant, il est clair qu'il ne parle pas chinois et qu'il ne fait que suivre une suite de règles prédéterminées. Cette expérience de pensée se veut une critique de l'IA symbolique et du test de Turing qui voudrait qu'une IA indiscernable d'un humain serait consciente ou, dans notre cas, qu'une IA produisant des réponses en chinois du même niveau que celles d'un sinophone comprendrait le chinois. Or ce n'est pas le cas : la limite de l'IA symbolique apparaît ici clairement. L'expérience de la chambre chinoise conduit Searle à conclure à l'impossibilité d'une « strong AI » douée d'une véritable compréhension qui semble réservée aux humains. En effet, à la différence des IA symboliques, nous possédons des systèmes sémantiques ancrés dans notre

¹ KIELA Douwe, BULAT Luana, VERO Anita L., et al., « Virtual embodiment: A scalable long-term strategy for artificial intelligence research », in *arXiv preprint arXiv:1610.07432*, 2016.

expérience sensori-motrice. Le chercheur Luc Steels, dans son article *Grounding Symbols through Evolutionary Language Games*¹, s'intéresse à l'acquisition de tels systèmes sémantiques, producteurs de sens, dont il distingue quatre composantes principales : monde, sensation, représentation et symbole. Le monde est premièrement perçu à travers une sensation qui affecte ensuite une représentation — un état biologique pour un organisme humain ou un état physique pour un ordinateur — correspondant à un symbole. Un système sémantique est donc « la relation entre le monde, la sensation, la représentation et le symbole »². Une fois qu'il existe une relation fonctionnelle entre ces quatre composantes, cela signifie que les représentations et les symboles sont ancrés et qu'un système sémantique a été appris. Mais le problème de l'ancrage des symboles ne s'arrête pas là. Lorsque les représentations sont ancrées, c'est-à-dire liées au monde et influencées par celui-ci, elles doivent encore provoquer une réponse motrice appropriée tandis que les symboles doivent être employés pour communiquer. Steels explique que la première partie du problème (l'acquisition d'un système sémantique par l'ancrage des représentations et des symboles) est beaucoup plus difficile que la seconde partie concernant le traitement des représentations et des symboles. Selon lui, c'est pour cette raison que la recherche en IA s'est principalement concentrée sur la partie traitement plutôt que sur la partie ancrage. Mais cet évitement du *grounding problem* n'est qu'un pis-aller, offrant une prise importante à des critiques comme celle de Searle. Le traitement symbolique s'est avéré efficace, notamment dans le cas des *LLMs*, mais il ne semble pas pouvoir mener seul à une compréhension profonde et au développement d'agents autonomes.

Pour pallier ce problème, des voix importantes, comme celles de Fei-Fei Li, s'élèvent en faveur d'une *embodied AI* conçue sur le modèle d'une intelligence humaine comprise d'abord comme une intelligence incorporée, fidèle à la description merleau-pontienne de l'expérience corporelle comme condition de toute perception et de toute action. Et, de fait, la mise au point de *world models* à l'instar de ceux qui poursuit *World Labs*, la start-up de Li, va dans le sens d'une IA en situation, toujours plongée dans un monde, tirant sa connaissance directement de perceptions multi-modales, voire de sensations, et non seulement de symboles. Cette conception nous renvoie à la seconde grande orientation de la recherche en IA qui entretient, depuis ses origines, une relation nourrie avec le thème de l'incarnation. Dans la culture populaire ce thème n'est jamais loin non plus : l'IA est souvent douée d'un corps, représentée dans les films et les livres sous la forme de robots

¹ STEELS Luc, « Grounding Symbols through Evolutionary Language Games » in *Simulating the Evolution of Language*, Angelo Cangelosi et Domenico Parisi (éd.), London, Springer London, 2002, p. 211-226.

² *Ibid.*

qui sont bien plus visuels que de simples *chatbots*. Au cours des différentes périodes de recherche sur l'IA et des nombreuses influences, l'incarnation a été plus ou moins importante mais n'a jamais disparu. *SHRDLU*, un programme conçu en 1968 par Terry Winograd, auquel un utilisateur pouvait demander de déplacer virtuellement des « blocs » (cônes, sphères et cubes) dans un monde virtuel en 3D, a été l'une des principales réalisations de cette conception de l'IA impliquant la capacité de percevoir et d'agir dans un environnement. Il en va de même pour *Shakey*, conçu à la fin des années 60, la première intelligence artificielle robotique capable d'agir dans un environnement physique. Souvent opposée à l'IA symbolique, la thèse de l'*embodiment*

implique que la meilleure façon d'acquérir une connaissance sémantique de niveau humain est de faire apprendre aux machines par le biais de l'expérience (physique) : si nous voulons enseigner à un système la véritable signification de « se cogner contre un mur », il suffit de le faire se cogner contre des murs à répétition.¹

L'apprentissage des humains et des animaux se fait nécessairement par le corps, de manière continue et unifiée depuis la naissance. Les données acquises par les différents sens ne sont pas traitées séparément mais mises-en-commun pour enrichir en permanence l'expérience vécue de chaque être. À l'inverse, les modèles sont entraînés selon des algorithmes, des méthodes et des types de données différents, et cessent d'apprendre une fois leur entraînement terminé. On comprend alors l'intérêt d'une IA qui serait incorporée, multi-modale, apprenant sans cesse d'après un mélange de *data* provenant de ses multiples « sens ». Une telle intelligence artificielle permettrait de créer des systèmes évolutifs, s'adaptant constamment à leur environnement. C'est dans l'esprit de réconcilier et de dépasser les difficultés rencontrées par les deux grandes aspirations de la recherche en IA, symbolisme et *embodiment*, qu'est apparu le paradigme de l'agent, et plus particulièrement de l'agent incorporé : robots, avatars virtuels ou encore appareils connectés. Dans leur article *Embodied AI Agents: Modeling the World*², la chercheuse Pascale Fung et ses collègues distinguent au moins trois avantages des agents incorporés : d'abord leur capacité à interagir avec un environnement physique, ensuite une amélioration des interactions homme-machine, et enfin, un mode d'apprentissage proche du notre, ce qui pourrait nous en apprendre sur nous-même. C'est que

¹ D. Kiela, L. Bulat, A.L. Vero, et al., « Virtual embodiment: A scalable long-term strategy for artificial intelligence research », art cit., Traduction personnelle : « implies that the best way for acquiring human-level semantics is to have machines learn through (physical) experience: if we want to teach a system the true meaning of “bumping into a wall”, we simply have to have it bump into walls repeatedly »

² FUNG Pascale, BACHRACH Yoram, CELIKYILMAZ Asli, et al., « Embodied AI Agents: Modeling the World », in *arXiv preprint arXiv:2506.22355*, 2025.

les agents incorporés, à la différence des agents désincarnés comme les *chatbots*, présentent le potentiel « d'apprendre et de se développer d'une manière similaire aux humains, en tirant parti d'une expérience sensorielle riche qui reflète la nôtre »¹.

Convaincus de l'importance d'un développement incorporé et évolutif de l'IA, Fei-Fei Li et ses assistants ont présenté une nouvelle architecture d'entraînement, *Deep Evolutionary Reinforcement Learning (DERL)*, dans l'article *Embodied intelligence via learning and evolution*² paru en 2021. Avec le *DERL* il s'agit de créer des agents incorporés selon trois axes de complexité : environnement, morphologie et contrôle intelligent. Mais cette expérience présente une caractéristique particulière : elle est menée dans *MuJoCo*, un environnement virtuel 3D simulant la physique. Les chercheurs ont ainsi mis-au-point des créatures virtuelles qu'ils nomment « UNIMAL » dotés de corps modulaires symétriques composés de membres de même forme pour une très grande variété de morphologies. En testant des centaines de configurations selon divers environnements 3D (plats, avec du relief, avec obstacles, etc...), diverses tâches (les créatures doivent se déplacer, pousser un objet, s'échapper, éviter des obstacles...) et en faisant se « reproduire » entre eux les *UNIMALs* les plus performants, les chercheurs mettent en évidence la sélection et la possibilité d'entraîner une intelligence morphologique. Le nombre très élevé d'itérations nécessaires à la simulation des processus évolutifs est, bien sûr, hors de portée de toute expérience physique, sans parler du niveau technique qui devrait être atteint afin de reproduire un niveau satisfaisant d'incarnation avec des robots. Les environnements virtuels offrent alors une solution idéale. Grâce à *MuJoCo*, Fei-Fei Li et ses collègues ont pu tester un très grand nombre de morphologies et entraîner toutes ces variantes dans des environnements différents. Finalement, leur expérience constitue l'une des plus importantes simulations d'évolution morphologique en *reinforcement learning* : quatre mille morphologies testées avec cinq millions d'itérations d'apprentissage pour chaque morphologie — l'équivalent de cinq millions de vies qui se succèdent, ce qui serait absolument impossible à mettre en œuvre dans le monde physique. Partant de ce constat, les auteurs de l'article *Virtual Embodiment: A Scalable Long-Term Strategy for Artificial Intelligence Research*³, défendent l'incorporation virtuelle comme méthode de recherche. D'une

¹ *Ibid.*, Traduction personnelle : « the potential for embodied agents to learn and develop in a manner similar to humans, leveraging a rich sensory experience that mirrors our own »

² GUPTA Agrim, SAVARESE Silvio, GANGULI Surya, et al., « Embodied intelligence via learning and evolution », in *Nature Communications*, n° 1, vol. 12, 6 octobre 2021, p. 5721.

³ D. Kiela, L. Bulat, A.L. Vero, et al., « Virtual embodiment: A scalable long-term strategy for artificial intelligence research », art cit.

manière générale, le *virtual embodiment* présente de nombreux avantages. D'abord, une évolutivité et un développement par incrémentation : la complexité des environnements virtuels peut augmenter à mesure de l'apprentissage des agents. C'est le cas dans l'article de Li où les environnements se complexifient, intégrant du relief ou des obstacles. Ensuite, une faisabilité à long terme : suivant l'amélioration des moyens techniques, les mondes virtuels s'enrichiront, offrant des perspectives et des défis de recherche sans cesse renouvelés pour de nombreuses années. Puis une rapidité d'itération : les cycles d'apprentissage peuvent être très rapides pour une évolution des agents à grande vitesse. Encore une fois, le *DERL* de Li en est la preuve : cinq mille itérations, équivalent à cinq mille vies calculées en quelques heures, ce qui permet de simuler des dynamiques complexes se déroulant sur un temps très long comme les phénomènes évolutifs. Ensuite, pas de nécessité d'une implication humaine permanente : là où pour des robots physiques l'interaction nécessaire à l'apprentissage ne peut provenir que d'une interaction avec un humain, dans l'incorporation virtuelle, les interactions peuvent avoir lieu entre les agents virtuels eux-mêmes. Enfin, d'un point de vue éthique : les environnements virtuels agissent comme des bacs-à-sable isolés dans lesquels on peut mener toutes sortes d'expériences sans craindre de conséquences regrettables sur le monde extérieur.

Ces nombreux avantages des agents virtuels incorporés expliquent leur usage croissant dans des domaines où les interactions riches et empathiques sont nécessaires, notamment pour la thérapie mais également pour peupler des mondes virtuels :

Dans les environnements virtuels tels que Second Life, Horizon Worlds et Sansar, les VEA [Virtual Embodied Agents] peuvent servir de guides, de mentors ou même d'amis, offrant aux utilisateurs une expérience plus immersive et interactive. Les VEA peuvent également contribuer à enrichir l'environnement virtuel, par exemple sous la forme de NPCs interagissant entre eux. Ces agents peuvent être conçus pour exprimer des émotions, de l'empathie et une intelligence sociale, ce qui en fait un atout inestimable dans les communautés virtuelles.¹

Et de la même manière, cette capacité à générer des comportements presque humains favorise l'emploi des agents virtuels incorporés dans la création de personnages pour le cinéma et le jeu vidéo:

¹ P. Fung, Y. Bachrach, A. Celikyilmaz, et al., « Embodied AI Agents: Modeling the World », art cit., Traduction personnelle : « In virtual environments like Second Life, Horizon Worlds, and Sansar, VEAs can serve as guides, mentors, or even friends, providing users with a more immersive and interactive experience. VEAs can also help provide a richer background virtual environment, for example in the form of NPCs interacting with each other. These agents can be rendered to exhibit emotions, empathy, and social intelligence, making them an invaluable asset in virtual communities. »

Un autre domaine dans lequel les VEAs sont à l'étude est celui des avatars de studios IA. Ces avatars peuvent être utilisés dans la production cinématographique, télévisuelle et vidéoludique afin de créer des personnages plus réalistes et captivants. En intégrant les VEA au processus de production, les créateurs peuvent concevoir des personnages qui sont non seulement visuellement impressionnants, mais également dotés d'une intelligence émotionnelle, capables de transmettre des émotions complexes et des nuances subtiles.¹

Il est vrai que ces agents incorporés sont particulièrement doués pour susciter de l'empathie. *Youtube* regorge de vidéos d'entraînements dans lesquelles des avatars apprennent à marcher, à courir, à sauter ou encore à jouer au foot dans des environnements 3D temps réel simulés physiquement comme *MuJoCo* ou le moteur de jeu *Unity*. Dans une vidéo d'*Open AI* qui a fait plus de dix millions de vues, on peut voir de petites créatures virtuelles apprendre à jouer à cache-cache dans un environnement tridimensionnel². Deux équipes de deux avatars sont en compétition : une équipe de *hiders*, qui doivent se cacher, et une équipe de *seekers*, qui doivent trouver les autres. Les avatars apprennent peu à peu, à force de millions de parties, à se déplacer dans l'espace pour se cacher puis à utiliser des outils mis-à-disposition par les développeurs, par exemple des cubes ou des rampes, pour se construire des abris (pour les *hiders*) ou franchir les obstacles (pour les *seekers*). Le couplage du *reinforcement learning* et de la simulation physique donne lieu à des simulations fascinante : on a l'impression d'assister au développement d'entités vivantes, animales ou même humaines. L'apprentissage de la marche bipède constitue, en particulier, un signe troublant d'humanité. Dès lors, il n'est pas étonnant que de nombreux éditeurs de jeu, à l'instar d'*Ubisoft*, se mettent au développement d'agents virtuels incorporés, capables de percevoir et d'agir dans des environnements simulés. Des telles intelligences artificielles permettraient non seulement d'améliorer les animations, les mouvements des personnages, mais également leurs possibilités d'interaction en les dotant d'une profondeur psychologique. Les *NPCs* sont souvent moqués pour la pauvreté des échanges qu'ils procurent aux joueurs, leurs actions et dialogues stéréotypés et répétitifs. L'intégration d'agents virtuels incorporés dans les jeux vidéo et autres espaces virtuels pourrait alors permettre de les peupler, d'en faire de véritable mondes : des lieux d'inter-agentivité, voire même d'intersubjectivité à mesure que les agents gagneraient en autonomie.

¹ *Ibid.*, Traduction personnelle : « Another area where VEAs are being explored is in AI studio avatars. These avatars can be used in film, television, and video game production to create more realistic and engaging characters. By incorporating VEAs into the production process, creators can craft characters that are not only visually stunning but also emotionally intelligent, capable of conveying complex emotions and subtle nuances. »

² Voir la vidéo « Multi-Agent Hide and Seek » sur la chaîne *Youtube* d'*Open AI* : <https://youtu.be/kopoLzvh5jY>

Par-delà prise de vues réelles et animation

Avec les *world models* et les agents virtuels incorporés, nous n'avons jamais été aussi proches de la conception de mondes virtuels au sens plein. Pas de simples représentations ou reproductions ne s'attachant qu'à la surface des choses, mais de véritables simulations dont la couche qui nous est accessible est redoublée d'une profondeur ontologique. L'apparence de ces simulations est le fait d'une structure interne, autonome, qui fait d'elles de véritables systèmes, des totalités douées d'un mouvement propre et de dynamiques évolutives. Les agents virtuels qui les peuplent ne sont plus les coquilles vides qu'étaient les personnages 3D animés, pas plus que les environnements générés ne sont de simples décors. Qu'on se rende compte du chemin parcouru.

Nous sommes partis de la prise de vues réelles qui repose sur une perception dans le monde phénoménologique — horizon de la perception, totalité inachevée et lieu d'intersubjectivité. Ce monde favorise l'expérience de la création artistique telle qu'elle est décrite par Dewey : une relation dynamique, métastable, faite d'équilibres et de déséquilibres successifs. L'artiste plongé au milieu du monde est en tension permanente avec son environnement, chaque tension entraînant un déséquilibre réglé par un nouvel équilibre et ainsi de suite. C'est au cours de ces allers-retours permanents, durant lesquelles l'artiste et le monde se façonnent mutuellement, que naît l'œuvre d'art. Dans le cas du cinéma, art collectif, l'environnement du cinéaste est parcouru de nombreuses subjectivités, à commencer par les interprètes avec lesquels il doit composer. L'environnement, et chacune de ces subjectivités, sont autonomes les uns par rapport aux autres. C'est cette autonomie, cette préexistence du monde dans lequel il est plongé, qui permet au cinéaste de réaliser des prises de vues. Contrairement aux arts de l'imaginaire comme la peinture ou l'animation, le cinéma en prise de vues réelles repose sur la perception. Sa matière première c'est le monde extérieur, physique, spatial, indépendant de l'esprit du réalisateur : un monde qui constitue l'horizon de sa perception et, de ce fait, possède toujours un temps d'avance, est riche de surprises et d'accidents. Même si l'exercice de mise en scène, et tout processus artistique en général, consiste bien à informer le monde, chaque artiste y laissant une marque plus ou moins profonde, cette marque est toujours dépendante du fond sur lequel elle s'imprime.

Nous avons vu alors qu'après la révolution de la reproduction photographique qui, par son automatisme, sort la subjectivité humaine du processus de création de l'image (du moins en apparence) ; un bon siècle plus tard, l'informatique s'attaque à un nouveau type de reproduction du

monde, pas seulement en surface mais en profondeur : la simulation. Avec la synthèse d'image, en particulier celle opérée par le rendu 3D, il s'agit de reproduire le monde comme système et de faire dépendre l'image de cette structure. C'est une nouvelle manière de mettre la subjectivité de côté : avec le « rendu », l'image n'est pas créée directement par l'utilisateur mais produite par l'intermédiaire d'un programme. Nous nous sommes alors interrogés sur le sort, a priori menacé, de la perception dans un tel mode de production d'image. L'obstacle principal tient dans la non-existence d'un monde déjà-là dans lequel la perception pourrait s'exercer. En effet, là où le photographe est d'emblée plongé dans un monde qui le précède, l'utilisateur d'un moteur 3D doit d'abord concevoir de toutes pièces une nature artificielle à partir de laquelle le programme calculera des vues. Cette absence d'un fond préalable sur lequel exercer une perception semble placer la synthèse d'images du côté de l'imaginaire et justifier sa catégorisation en tant qu'animation. Pour autant, le monde virtuel 3D, s'il est créé par un infographiste, ne dépend pas de lui pour persister dans l'existence. Nous avons vu que les scènes 3D étaient au moins semi-persistantes : à défaut d'exister en permanence sous une forme visuelle, elles existent au moins comme descriptions mathématiques et numériques. Surtout, elles rendent possible l'expérience d'un espace tridimensionnel dans lequel se déplacer et contrôler la caméra virtuelle qui fait office de point de vue. Dès lors, les moteurs de rendu 3D, et à plus forte raison, les moteurs de jeu vidéo en temps réel permettent bien une perception instrumentée, prolongation de la subjectivité de l'utilisateur, et le déploiement de la dynamique du corps propre et du corps objectif dans l'espace virtuel. Le premier obstacle au concept de la « prise de vues virtuelles » est alors dépassé, d'autant plus que tout est conçu, dans les moteurs de rendu, par analogie à la prise de vues réelles : caméra dont les paramètres optiques sont calqués sur les caméras physiques, nécessité de placer des spots virtuels pour éclairer les scènes 3D... Finalement, la création d'images de synthèse 3D s'apparente bien à une forme de prise de vues dans un espace véritable, bien que virtuel, où l'on peut se déplacer, orienter la caméra et cadrer, pour ensuite appuyer sur un déclencheur numérique et laisser place à un processus automatique de capture d'une vue.

Mais c'est alors que nous avons rencontré à un second obstacle, plus redoutable encore : l'automatisme. Ce qui caractérise l'expérience du monde phénoménologique c'est son inachèvement, son apparaître constant qui n'est jamais épuisé par la perception. Or, les simulations 3D se heurtent toujours à leur manque d'autonomie et leur automatisme stupide. Elles ne sauraient produire de réelle surprise, dans le sens où tout ce qui existe dans leurs espaces virtuels, a été décidé et décrit explicitement par leur concepteur. Les jeux vidéo, et par extension les machinimas, s'en

sortent un peu mieux que les scènes 3D, dans lesquelles travaillent les animateurs, parce qu'ils offrent aux joueurs un univers déjà constitué, un semblant de totalité englobante dans laquelle se plonger. Mais nous l'avons vu avec l'exemple de jeux procéduraux : aussi complexes soient-ils, ces systèmes n'entretiennent l'illusion qu'un temps, et le voile des apparences finit toujours par se déchirer, révélant des abîmes. Mathieu Triclot évoquait à ce propos la période préliminaire à la phase de jeu durant laquelle il faut cartographier les limites du programme — voir quelle porte s'ouvre, déterminer les contours de la *map*, etc... — pour s'assurer ensuite de rester dans ses frontières. Les simulations 3D appartiennent ainsi à la mauvaise automaticité, l'automaticité des machines qu'on oppose souvent à la créativité du vivant. Elles se contentent d'agencer des éléments préfabriqués dont elles viennent tôt ou tard à bout des combinaisons. La perception est permise mais ces scènes virtuelles ne sont ni des totalités autonomes, ni des apparaîtres constants, encore moins des lieux d'intersubjectivité. En un mot : la vie manque terriblement à ces espaces qui ne sont pas des mondes. Pourtant, l'automatisme stupide des automates n'est pas une fatalité. Dans son livre *Métamorphoses de l'intelligence*¹, Catherine Malabou rappelle que l'automatisme n'est pas opposé au vivant. Bien au contraire, le vivant fait sans cesse appel à l'automatisation, elle est inévitable et fait partie du processus d'apprentissage. Ainsi l'opposition entre intelligence et automatisme est moins importante qu'il n'y paraît au premier abord, c'est « une tension interne à l'intelligence et à l'automatisme eux-mêmes. »². Malabou s'appuie notamment sur les travaux de Dewey qui défend la part d'automatisme de l'intelligence. Pour lui, « tous les actes de l'intelligence sont instrumentaux »³ dans le sens où ils ont une utilité pratique pour la résolution de problèmes. Et cette résolution passe par une automatisation : résoudre un problème, c'est le comprendre et, par extension, transformer la situation et la réaction en automatismes. Ainsi, « résoudre revient toujours d'abord à mettre en mouvement, et mettre en mouvement revient toujours d'une manière ou d'une autre, à automatiser. »⁴. Nous avons évoqué en introduction les deux sens contraires au terme « automatisme » que distingue Malabou : d'une part, ce qui désigne un « mouvement involontaire, qui n'a pas en soi-même son "âme" », et, d'autre part, « ce qui se meut de soi-même, spontanément »⁵. Toute la pensée de la technique cherche donc à comprendre les rapports entre les

¹ C. Malabou, *Métamorphoses de l'intelligence*, *op. cit.*

² *Ibid.*, p. 126.

³ *Ibid.*

⁴ *Ibid.*, p. 127.

⁵ *Ibid.*

deux appréciations de l'automatisme. Finalement, selon Dewey, l'intelligence est ce qui permet de lutter contre le premier automatisme au moyen du second, « ce qui justement oppose le pouvoir de l'automatisme aux automatismes du pouvoir »¹. Cette double conception a constitué, dès les débuts de la recherche en IA, deux pôles théoriques opposés :

Premier pôle — certains informaticiens ont considéré que le cerveau n'était qu'une machine à apprendre. Une telle machine, pensaient-ils, pouvait être dupliquée sans problème par des ordinateurs eux-mêmes conçus comme de simples machines à calculer. De l'autre — second pôle — d'autres chercheurs ont immédiatement reconnu l'importance que pouvait avoir la plasticité cérébrale pour le développement de l'IA.²

Les précurseurs de la cybernétique comme Turing, Ashby ou Von Neumann, se sont intéressés à la plasticité dans le but d'inventer « une machine capable de s'interrompre pour mieux se réorganiser elle-même, de dialectiser son automatisme, de suspendre le processus de la répétition à l'identique. »³. Cette position contre les automates et en faveur de l'indétermination des machines apparaît clairement chez Gilbert Simondon, fortement influencé par la cybernétique :

Le véritable perfectionnement des machines, celui dont on peut dire qu'il élève le degré de technicité, correspond non pas à un accroissement de l'automatisme, mais au contraire au fait que le fonctionnement d'une machine recèle une certaine marge d'indétermination. C'est cette marge qui permet à une machine d'être sensible à une information extérieure.⁴

Il s'agit de concevoir des machines douées d'une marge d'indétermination, d'une forme de liberté plus volontiers attribuée au vivant. C'est là que réside la véritable perfection, non dans la surenchère d'automatismes, à l'instar d'une superproduction comme *Red Dead Redemption II*, mais dans une programmation ouverte qui laisse un espace pour l'imprévisible et même une forme de plasticité artificielle. Or, nous avons vu que cette plasticité artificielle est devenue une réalité avec l'avènement du *deep learning*. Enfin, l'IA générative et les modèles de diffusion permettent la dimension de surprise qui caractérise l'apparaître permanent du monde. L'autonomie incroyable de

¹ *Ibid.*

² *Ibid.*, p. 145.

³ *Ibid.*, p. 146.

⁴ SIMONDON Gilbert, *Du mode d'existence des objets techniques*, Paris, Aubier, 2012, p. 11.

ces nouvelles machines est inédite dans l'histoire. Il est possible de diriger des modèles, comme on dirige des interprètes, pour générer des images et des séquences animées. Le *machine learning* libère les modèles du joug de l'automatisme bête, « mouvement involontaire » pour introduire de l'indétermination, de la plasticité : l'automatisme du vivant qui est un mouvement interne et spontané.

Pourtant, nous avons vu au chapitre 6 que l'IA générative et les modèles de diffusion ne constituaient pas une solution ultime aux limites rencontrées par la synthèse 3D. Bien que prometteurs quant au problème de l'automatisme, les modèles génératifs font moins bien que les moteurs de rendu en termes de persistance spatiale. Ils n'offrent qu'une perception très pauvre voire inexistante, plus proche du rêve, qui les éloigne du paradigme de la prise de vues pour les ramener dans le giron de la création imaginaire et de l'animation. Mais, derrière ce semi-échec, on distingue déjà les contours d'une solution qui dépasse les espérances. Les différentes avancées de l'IA générative, *LLMs* puis modèles de génération d'images et de vidéos, sont des étapes sur le chemin qui mène aux *world models*, la génération de mondes. Comme nous l'avons vu, ces modèles permettent de générer des mondes consistants, des lieux autonomes dans lesquels se déplacer, percevoir et agir, toujours en train d'apparaître et même peuplé d'agents virtuels dont l'autonomie, on peut l'anticiper, va aller en augmentant jusqu'à faire de ces mondes des lieux d'intersubjectivité, ou au moins d'interagentivité. L'adoption d'un nouveau paradigme pour l'intelligence artificielle — passage d'une intelligence symbolique, désincarnée à une intelligence encorporee — promet de déployer pleinement cet automatisme intelligent qui est d'abord celui du corps avant d'être celui de l'esprit. L'ordinateur est en passe d'accomplir sa promesse d'être bien plus qu'un automate, un être technique doué d'une marge d'indétermination qui le rend plastique, un « autonome »¹. Enfin, nous entrevoyons la possibilité de véritables mondes virtuels cochant tous les critères phénoménologiques, et, avec eux, l'avènement d'une prise de vues virtuelles dans son sens le plus fort : tournage dans des simulations « dont la profondeur et l'indétermination répondent à l'insatisfaction qui caractérise la vie »².

Dans son livre *Automates, robots et humains virtuels dans les arts vivants*³, le théoricien des arts numériques, Edmond Couchot, dresse un historique de l'utilisation des

¹ Terme du théoricien des arts numériques Edmond Couchot

² R. Barbaras, *La perception. Essai sur le sensible*, op. cit., p. 54.

³ COUCHOT Edmond, *Automates, robots et humains virtuels dans les arts vivants*, Saint-Denis, Presses universitaires de Vincennes, « Théâtres du monde », 2022, 220 p.

automates dans les arts de la scène et montre que ceux-ci ont toujours entretenu une relation étroite. Mais pour la première fois de l'histoire, les créatures artificielles se retrouvent douées de cette plasticité qui les éloigne des automates pour les rapprocher des vivants. L'automatisme des automates, qui est aussi celui de l'animation, laisse place à l'intelligence des automatismes. Du côté virtuel, les *world models* offrent une flexibilité et des possibilités immenses pour la création de mondes, tandis que les agents incorporés annoncent l'abandon des marionnettes 3D, et leur animation laborieuse, au profit de personnages virtuels doués d'une profondeur psychologique. Au point qu'on pourrait imaginer l'émergence d'un mode de réalisation qui reposerait sur la direction d'acteurs virtuels aux personnalités tout aussi riches, aux voix toutes aussi incarnées, pour une expérience tout aussi profonde que le tournage physique avec des interprètes de chair et d'os. Se référant aux techniques déjà en place dans le théâtre et la danse contemporaine qui incluent des robots et des créatures virtuelles, Couchot parle d' « arts vivants artificiels » pour marquer ce tournant vers la plasticité artificielle. Et de la même manière, ces mondes et ces acteurs virtuels qui, on peut en faire le pari, iront vers toujours plus d'autonomie, promettent une prise de vues virtuelles au sens fort du terme, un *artificial live action* qui pourrait bien constituer un nouveau cinéma hors des catégories de prise de vues réelles et d'animation.

Études d'œuvres

Nikita Diakur et la simulation du mouvement

Parmi les représentants possibles de cette nouvelle prise de vues virtuelles, le réalisateur Nikita Diakur aurait toute sa place. Bien que classé dans la catégorie de l'animation 3D, Diakur aime à parler d'« animation interactive » pour faire ressortir la particularité de ses œuvres bien différentes des canons du genre et des productions de *Pixar*. En effet, plutôt que de concevoir des « poses clés » et de laisser le logiciel interpoler entre celles-ci, comme c'est le cas dans l'animation *keyframe*, Diakur crée des marionnettes dont les articulations sont reliées à des fils virtuels et dont les corps sont simulés physiquement. Cette méthode d'animation permet une grande part d'aléatoire car le contrôle de l'animateur est très sommaire : en marionnettiste, celui-ci se contente d'agiter les fils pour infliger des secousses à ses personnages dont le mouvement est alors délégué au moteur physique du logiciel 3D qui se charge de simuler la gravité, les chocs et l'ensemble des forces qui s'exercent sur les corps virtuels. Cette méthode d'animation que Diakur qualifie parfois de « digital puppetry » offre des possibilités nouvelles, très originales de mouvement. Là où l'animation *keyframe* nous avait habitué à des mouvements lisses, parfaits, qui ne heurtent pas la sensibilité des spectateurs, ceux des personnages de Diakur sont saccadés voire *buggés*. Car la dimension interactive de son travail passe aussi par l'intégration des *bugs* et des artefacts dont l'animateur se délecte. Dans ses films, les personnages se déplacent de manière gauche, peinent à réaliser des tâches simples, chutent en permanence, se cognent entre eux et aux éléments du décors qui s'écroulent parfois. Le thème de la chute qui revient constamment dans ses films est particulièrement révélateur du traitement du mouvement par Diakur. Le court métrage *Fest* met en scène une *rave party* déchaînée au cours de laquelle l'un des personnages saute à l'élastique en heurtant violemment l'ensemble des éléments se trouvant sur sa trajectoire, dans son trailer *Final Flight* réalisé pour le festival d'animation de Viborg quatre personnages désarticulés effectuent un saut en chute libre depuis un avion, dans son documentaire *Backflip* une doublure 3D de lui-même apprend à réaliser un salto arrière en chutant une bonne centaine de fois consécutives, tandis que son installation *Downonly* expose un ordinateur calculant sans cesse la chute de personnages 3D dans un environnement virtuel et se termine par la chute dans, le lieu d'exposition, de l'ordinateur lui-même. C'est que la chute constitue un exemple pur de mouvement dysfonctionnant qui, dès lors qu'il ne peut plus s'effacer derrière l'utilité ou l'habitude, devient évident, mouvement pour lui-même. D'ailleurs, il n'est pas surprenant que Diakur tire souvent son inspiration des *video gags*.

Dans l'absurdité de ces séquences, le mouvement — qui est souvent celui d'une chute — est brutalement extrait du flot continu de l'automatisme monotone pour apparaître dans toute son étrangeté et sa singularité : le mouvement de ce jeune homme qui fait tourner à toute vitesse le tourniquet d'un jardin d'enfant en y appliquant la roue avant de sa moto couchée sur le côté, ou celui de cette femme d'une banlieue russe qui saute à l'élastique au milieu des barres de HLM soviétiques, rompent soudain avec la normalité des mouvements quotidiens dont ils s'émancipent, ce sont des mouvements autonomes produits d'une activité interne. Avec leur gaucherie et leur lourdeur, leurs comportements absurdes et leur assujettissement à la physique (ce qui n'est pas le cas des personnages animés de manière classique), les personnages de Diakur retrouvent un corps, une corporalité qu'avaient perdu les personnages 3D. Et leur caractère désarticulé de pantins parvient paradoxalement à provoquer une grande empathie : on éprouve de la peur, de l'amusement, de la joie ou de la tristesse pour eux. Bien souvent, on les trouve stupide mais jamais complètement vides ; on croit distinguer dans les mouvements qui les traversent sans cesse les marques d'une dynamique interne et autonome. Surtout, ils portent en eux la métastabilité du vivant : là où un personnage 3D animé en *keyframe* entre dans une immobilité parfaite dès lors qu'il n'est pas sollicité par un animateur, les personnages de Diakur sont sans cesse soumis aux dynamiques du moteur physique, tremblants, en tension, toujours susceptibles de se heurter à un autre corps ou simplement de s'écrouler sur eux-mêmes : toujours au bord de la chute. Cette démarche artistique de Diakur qui poursuit le mouvement pour lui-même (et non comme moyen) afin de faire réapparaître le corps n'est pas sans rappeler l'ambition de nombreux chorégraphes et de la danse en général. Comme l'écrit Edmond Couchot,

C'est dans la danse que l'expression corporelle paraît, en son fondement, la plus intense et la plus corporelle, et que les mouvements du corps ne renvoient qu'à leurs propres formes. [...] La danse est une sorte de musique des corps dont les éléments constitutifs, structurés selon des règles ou des conventions comme ceux de la musique, ne représentent qu'eux-mêmes.¹

La danse est une forme de mouvement particulièrement importante, sans doute la plus importante en ce qu'elle constitue une marque profonde d'humanité : si l'ensemble des mouvements quotidiens ont une dimension utilitaire qui peut jeter le doute sur l'autonomie de celui qui les réalise — de nombreux mouvements qu'on fait sans vraiment réfléchir peuvent apparaître sous le coup d'un automatisme extérieur —, nul doute n'est permis dans le cas de la danse, ces mouvements sont si

¹ Edmond Couchot, *Automates, robots et humains virtuels dans les arts vivants*, Théâtres du monde (Presses universitaires de Vincennes, 2022), p.121

étranges, si détachés du cours normal qu'ils ressemblent bien à ceux d'une conscience qui reprendrait le contrôle de son corps, de l'intérieur. La danse vise autre chose que les gestes fonctionnels ou sociaux du quotidien, elle vise le mouvement pour le mouvement dans le but de produire une expérience esthétique ; elle dévoile des possibilités insoupçonnées du corps humain. Celle-ci constitue alors un défi d'autant plus important pour les créatures virtuelles et les robots, plus habitués aux mouvements utiles, que l'enjeu est de taille : faire danser un robot, c'est le doter d'une intériorité. Dans son livre, Couchot évoque plusieurs de ces tentatives, notamment le spectacle *Robot!* de la chorégraphe Blanca Li qui s'interroge sur la manière de créer de l'empathie pour les robots qu'elle met en scène. Finalement, c'est leur maladresse d'automates qu'elle utilise :

C'est leur côté enfantin, fait de fragilité, et leur gaucherie, que j'ai utilisées dans leurs rapports avec les danseurs, pour rendre sensibles ces petits robots humanoïdes. Pour la première fois, l'idée que ceux-ci puissent devenir des compagnons aimables et attachants devient évidente pour le grand public.¹

On comprend alors mieux l'empathie étrange que l'on éprouve pour les créatures désarticulées de Nikita Diakur. C'est que derrière leur maladresse, et les tressautements causés par les bugs informatiques du système se trouve la beauté du mouvement imprévisible, celle de l'automatisme intelligent du vivant. Et de fait, dans son documentaire *Backflip*, lorsqu'on voit son avatar s'entraîner à faire des saltos arrière et tomber, tomber, encore tomber pour finalement se tenir droit, on éprouve un immense sentiment d'humanité, sans doute plus impressionnant et surtout plus profond, plus émouvant que n'importe quel dialogue avec un LLM.

Et dans sa poursuite du mouvement pour lui-même, Diakur cherche sans cesse à complexifier les automatismes de ses personnages, et de leur environnement, pour les rendre les plus complets et les plus profonds possibles : il cherche à construire des systèmes. Dans le logiciel qu'il utilise, les simulations physiques sont désignées sous le terme de « dynamics ». Tout l'enjeu pour Diakur c'est détendre au maximum l'emploi de ces *dynamics*, de créer le plus possible de dynamiques entre les différents éléments de la scène. Et en effet, plus ces systèmes physiques contrôlent d'éléments dans une scène, plus la simulation est riche et plus l'on s'éloigne de l'animation et de ses objets vides, sans mouvement interne. Ce sont particulièrement les personnages qui gagnent en profondeur : s'ils ne sont pas véritablement doués d'autonomie, ils ont une agentivité bien plus forte que les personnages 3D des films d'animation classiques qui sont

¹ Bianca Li, citée par Edmond Couchot, *Automates, robots et humains virtuels dans les arts vivants*, coll. « Théâtres du monde », Paris, Presses universitaires de Vincennes, 2022, p. 128.

souvent de simples enveloppes. Diakur parle même à leur propos d' « acteurs » ou d' « amis virtuels ». Ils prennent part à des systèmes dans lesquels ils sont soumis à des forces, mais sur lesquels ils exercent également une influence. Lorsqu'un personnage en heurte un autre, par exemple, il y a un véritable choc qui envoie, selon la force de l'impact valser son camarade plus ou moins loin. Dans l'une de ses expériences, Diakur a mis au point une simulation de machine de fête foraine pour attraper des peluches avec une pince contrôlée par une manette. Dans un film d'animation classique, un animateur aurait animé de manière distincte la main d'un personnage en train de pousser la manette, puis le mouvement de la manette, et enfin celui de la pince en train de saisir la peluche afin de donner l'illusion d'un lien causal entre ces trois mouvements. Nulle illusion avec la simulation de Diakur : celui-ci se contente d'animer le premier mouvement de la chaîne causale — la main poussant la manette — et laisse les *dynamics* faire le reste : la pince s'active et saisit un lot de la machine sous l'effet d'un mécanisme. D'une manière générale, pour réaliser toutes sortes d'effets et d'animations Diakur privilégie les systèmes qui sont comme de micro jeux-video. Pour animer le vol des feuilles mortes, Diakur place un ventilateur virtuel dans la scène dont il peut contrôler la puissance du souffle, pour mettre en mouvement une voiture, Diakur branche une manette de *Playstation* dont le *joystick* lui permet de contrôler la vitesse et la trajectoire du véhicule. C'est la mise en place de ces systèmes complets et fonctionnels dans ses scènes 3D qui permet à Diakur de parler d'animation interactive. Prenant pour référence le réalisateur David O'Reilly et son film-jeu *Everything* qui met en scène l'ensemble de l'univers dont le joueur peut visiter les différentes échelles, depuis le monde microscopique des cellules jusqu'à l'immensité des galaxies, le machinima, déjà cité, *Martin Pleure* de Jonathan Pinel tourné dans *GTA* ou encore les vidéos d'entraînement d'avatars 3D de *Google AI*, la définition que donne Diakur de l'animation interactive est en réalité très proche de celle que nous avons donné de la prise de vues virtuelles : prise de vues dans des mondes virtuelles. L'analogie du tournage est d'autant plus évidente que Diakur décide d'inclure dans son esthétique les dessous techniques de la simulation : *rigs* apparents, maillage de la géométrie 3D visible, caméras de rendu, *bugs* intégrés au montage final qui deviennent partie prenante de la narration, *camera mapping* pour les décors et les vêtements. Choissant de montrer tout ce qui est habituellement caché, Diakur assume le médium de la 3D prenant le contre pied de la tradition principale de la *CGI* qui cherche sans cesse à entretenir la confusion entre ses images et les images filmées, ou du moins à faire oublier que ses images ont été générées au cœur d'un ordinateur. Diakur assume sa position jusqu'au bout en se proclamant même, assez ironiquement, « ugly filmmaker » — réalisateurs de laideurs —, en cherchant des designs volontairement géométriques, disproportionnés et buggés pour ses personnages et ses décors de

manière à révéler la nature synthétique de ses images. Cette laideur — le terme à une vocation provocatrice car en réalité les films de Diakur sont assez beaux — revendiquée, couplée à l'animation interactive qui, ensemble, constituent l'identité artistique de Diakur sont particulièrement évidentes dans son court-métrage-manifeste *Ugly*. Mettant en scène le destin malheureux d'un chat persécuté pour sa laideur, le film déploie tout un environnement virtuel dont Diakur assume pleinement la dimension simulationnelle. Le réalisateur tourne dans une simulation informatique et veut le faire savoir. Au point qu'au cours du film, Diakur fait brièvement passer le spectateur de l'autre côté, dans les coulisses : on se retrouve dans le *viewport* du logiciel, tous les personnages et les éléments de décors flottant dans un chaos numérique sur un fond noir. On aperçoit même le cameraman portant la caméra virtuelle à partir de laquelle se fait le rendu des images de synthèse. Et en effet, de nombreuses scènes du film sont tournées avec un opérateur de caméra : concrètement, un personnage simulé de la même manière que les autres et portant la caméra virtuelle. Il n'apparaît jamais hormis dans cette scène de chaos flottant. En revanche, on peut apercevoir son ombre dans de nombreuses autres scènes, preuve que *Ugly* a bien été tourné. Et il en va de même des plans aériens qui ont été capturés par un drone simulé physiquement et contrôlé par un opérateur au sol — l'un des personnages. En contemplant cette petite oasis de vie artificielle au milieu d'un néant numérique, on prend alors conscience de la fragilité de ce monde virtuel, de ses personnages et de ce film. Si Diakur a mis près de quatre ans pour réaliser ce court-métrage d'une dizaine de minutes, c'est parce qu'animer ses personnages était « comme travailler avec des êtres vivants ». Et comme avec de « vrais » acteurs, Diakur fait des prises : parfois il doit relancer une simulation cinquante fois avant d'avoir un résultat satisfaisant. Cette méthode ouvre la voie à des surprises de la part de l'ordinateur qui se retrouve presque au rang de collaborateur artistique.

Cette collaboration s'accroît encore avec la réalisation du documentaire *Backflip* qui met en scène une doublure virtuelle de Diakur entraînée, par le biais du *machine learning*, à réaliser un salto arrière dans un environnement 3D simulé physiquement :

Tenter un salto arrière n'est pas sans danger. On peut se briser la nuque, atterrir sur la tête ou mal retomber sur les poignets. Rien de tout cela n'est agréable, alors c'est mon avatar qui s'en charge. Il s'entraîne sur un processeur à six cœurs à l'aide du Machine Learning. Le processeur n'est pas le plus récent, mais il parvient tout de même à calculer six sauts par itération. Une itération prend une

minute, ce qui représente 360 sauts par heure et 8 640 sauts par jour. Je ne serais pas capable de sauter autant moi-même.¹

Au delà, de l'apprentissage des mouvements corporels de l'avatar, l'intelligence artificielle est également utilisée pour générer le son — dialogue et musique — ainsi que le *lip sync* — la synchronisation des lèvres avec les paroles — renforçant encore l'impression d'avoir affaire à un être artificiel autonome. Pour Diakur, qui disait déjà travailler avec des acteurs virtuels ou même des êtres vivants, ce projet représente un nouveau pas vers le mouvement autonome et la prise de vues virtuelles. Au lieu de simuler le comportement physique de marionnettes reliées à des fils virtuels, le réalisateur s'affranchit de tout contrôle externe. Si la méthode d'animation de la *digital puppetry* lui permettait d'obtenir des résultats intéressants, il n'en demeurerait pas moins que la marionnette restait du côté des automates, un être inanimé. Avec le *machine learning*, il s'agit de remplacer la marionnette par un avatar dont le mouvement est d'origine interne. C'est un pas hors de l'animation, en témoigne la description technique du film qui n'est pas classé en tant qu'« animation » mais en tant que « simulation ».

Diakur dit avoir d'abord été inspiré par des vidéos d'entraînements d'agents virtuels incorporés, notamment celles de *Deepmind* montrant des avatars qui apprennent à marcher tout seul dans un environnement physique simulé. D'une manière plus précise, l'algorithme utilisé par Diakur pour *Backflip* se base sur l'article scientifique *DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills*² paru en 2018. Le modèle *DeepMimic* a ceci de particulier qu'il combine à la fois entraînement à partir d'une base de données de mouvements — vidéos ou séquences d'animation en 3D — et *reinforcement learning (RL)* — la réalisation de tâches spécifiques dont le résultat donne lieu à des *feedback* positifs ou négatifs. L'article part du constat que les simulations physiques sont largement utilisées dans la création d'image de synthèse pour des phénomènes passifs (vent, liquides, feu...) mais rarement pour les phénomènes actifs, *e.g.* le mouvement autonome d'origine interne des êtres vivants. Dans ce cas, l'animation *keyframe* est privilégiée puisqu'elle offre des résultats bien supérieurs et surtout un meilleur contrôle. Dans cette poursuite de la simulation physique, qui permettrait à des animations

¹ Ma traduction de : « Attempting a backflip is not safe. You can break your neck, or land on your head, or land badly on your wrists. None of that is nice, so my avatar does the trick. It practices on a 6-core processor with the help of Machine Learning. The processor is not the newest but still calculates 6 jumps per iteration. One iteration takes one minute, this is 360 jumps in an hour and 8.640 jumps in a day. I wouldn't be able to jump so much myself. », Nikita Diakur, « Synopsis », dossier de presse de *Backflip* [PDF en ligne], 2022, disponible sur : https://backflip.training/press/backflip_info.pdf, consulté le 04/09/2026

² Xue Bin Peng et al., « DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills », *ACM Transactions on Graphics* 37, n° 4 (2018): 1-14, <https://doi.org/10.1145/3197517.3201311>.

de s'adapter à une multitude d'environnements¹, le *reinforcement learning* réalisé dans un environnement physique offre de belles perspectives. Mais cette technique présente une difficulté importante : elle produit souvent des artefacts et des animations tremblantes tout en offrant peu de possibilités de contrôle. En repartant de travaux précédents, les auteurs de l'article entendent dépasser ces obstacles en couplant les simulations physique du *RL* à des données d'animation :

La contribution centrale de notre article est un cadre d'animation de personnages basée sur la physique, qui combine l'apprentissage par renforcement orienté objectif avec des données pouvant être fournies sous forme de séquences de capture de mouvement ou d'animations par images clés. [...] Notre système reçoit en entrée un modèle de personnage, un ensemble correspondant de mouvements de référence cinématiques, et une tâche définie par une fonction de récompense. Il synthétise alors un contrôleur permettant au personnage d'imiter les mouvements de référence, tout en satisfaisant les objectifs de la tâche, tels que frapper une cible ou courir dans une direction donnée sur un terrain irrégulier. [...] Le résultat final de notre système est une politique qui permet à un personnage simulé d'imiter les comportements des mouvements de référence tout en remplissant les objectifs de tâche spécifiés.²

Pour mener à bien leur expérience, les auteurs de *Deep Mimic* ont ainsi entraîné différents avatars avec une vingtaine de mouvement spécifiques, notamment le salto arrière. Le code du modèle étant *open source*, Diakur a pu s'en servir pour entraîner son propre avatar dans l'espace 3D simulé physiquement du moteur de jeu *Unity*. Après avoir laissé tourner le modèle pendant quelques heures sur des tâches préliminaires comme l'apprentissage de la position debout et puis de la roue, l'avatar de Diakur a pu se lancer à l'assaut du *backflip*. Finalement, il aura fallu une journée entière de calcul et plusieurs milliers de sauts pour que l'avatar parviennent enfin à retomber sur ses pieds après avoir réalisé un salto arrière. Diakur présente ce projet comme une expérience personnelle née d'une fascination et d'une peur vis-à-vis de la technologie qui atteint son paroxysme dans la création d'une doublure numérique qui, d'une certaine manière, étend l'identité de son modèle original :

¹ Particulièrement utile pour le jeu vidéo où les personnages doivent accomplir des très nombreuses tâches dans des environnements variées. La mise au point d'animations adaptatives permettrait ainsi d'économiser le travail des animateurs tout en permettant une plus grande variété d'action, imaginons par exemple un cycle de marche qui pourrait s'adapter au relief de n'importe quel terrain.

² Ma traduction de : « The central contribution of our paper is a framework for physics-based character animation that combines goal-directed reinforcement learning with data, which may be provided in the form of motion capture clips or keyframed animations. [...] Our system receives as input a character model, a corresponding set of kinematic reference motions, and a task defined by a reward function. It then synthesizes a controller that enables the character to imitate the reference motions, while also satisfying task objectives, such as striking a target or running in a desired direction over irregular terrain. [...] The final result of our system is a policy that enables a simulated character to imitate the behaviours from the reference motions while also fulfilling the specified task objectives. », Xue Bin Peng et al., « DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills », *op. cit.*

C'est pourquoi j'ai construit mon avatar. D'une certaine manière, c'est ma façon de faire face à l'expérience que je viens de décrire, en apprenant virtuellement le salto arrière par l'intermédiaire d'une copie de moi-même. L'avatar ne ressent aucune douleur et continuerait, quoi qu'il arrive.¹

Le documentaire a eu un grand succès, remportant des prix dans de nombreux festivals et touchant un large public. De fait, il constitue une introduction fascinante, drôle, émouvante et parfois même effrayante à l'intelligence artificielle et au *deep learning*. Il y a quelque chose de très doux et de très fragile dans cet avatar trébuchant qui apparaît doué d'une véritable intériorité, et l'on ne peut s'empêcher de compatir à ses chute autant qu'à sa réussite, comme si la réalisation de ce salto arrière parachevant le processus d'apprentissage l'avait rendu vivant. *Backflip* constitue sans doute l'un de tous premiers exemples de prises de vues virtuelles et — je dois lui rendre ce qui lui appartient — ma première inspiration, ma première intuition quant à ce concept.

Ian Cheng et la simulation de mondes

Dans la continuité du travail de Nikita Diakur, l'artiste américain Ian Cheng, qui travaille avec des simulations 3D dans le moteur de jeu *Unity*, poursuit, lui aussi, l'automatisme intelligent du vivant. Dans sa quête holistique de l'émergence — l'idée qu'un système complexe peut donner lieu à des propriétés et des comportements qui ne sont pas présents dans ses parties — Cheng cherche à concevoir des systèmes et, plus encore, des mondes, ce qui l'amène à qualifier sa pratique artistique de « worlding » ou, en français, de « mondiation ». Tout comme Diakur, Cheng commence par formuler le souhait d'une méthode d'animation plus organique et plus flexible que l'animation 3D traditionnelle. Après un court passage dans l'industrie des effets spéciaux, l'artiste, prenant le contre-pied de cette expérience professionnelle, éprouve le besoin d'échapper au perfectionnisme inhérent à l'animation. Une question s'impose alors : « comment une animation pourrait-elle s'animer elle-même » ?². La découverte du moteur de rendu en temps réel *Unity* lui apparaît comme une réponse : conçu pour le jeu vidéo, cet outil facilite l'automatisation et la mise en place de systèmes. Il commence à jouer avec les simulations physiques du moteur, en particulier les « ballspring physics » — un système de boules solides reliées par des ressorts, permettant de

¹ Ma traduction de : « That is why I built my avatar. In a way, it is me coping with the above experience by virtually learning the backflip via a copy of myself. The avatar doesn't feel any pain and would keep on going, no matter what. », Nikita Diakur, « Director Statement », dossier de presse de *Backflip* [PDF en ligne], 2022, disponible sur : https://backflip.training/press/backflip_info.pdf, consulté le 04/09/2026.

² Ma traduction de : « how an animation might be able to animate itself ? », Ian Cheng, *Ian Cheng DMA Lecture Series 2017-18*, conférence à UCLA Design Media Arts, 7 novembre 2017, UCLA Design Media Arts, vidéo en ligne, 01:16:56, URL : <https://youtu.be/gBOfkWTUnvI>, consultée le 4 septembre 2026.

concevoir toutes sortes structures dynamiques et évolutives. Cheng s’empare de cet outil pour animer des organismes au design fluide, mou et flexible, produisant des résultats rares dans l’animations 3D. Dans certains de ses tests, Cheng va même jusqu’à coupler ces organismes à des réseaux de neurones pour contrôler leurs mouvements et les faire participer à des jeux évolutifs, comme l’évasion d’un piège virtuel à l’instar des simulations d’*Open AI* déjà évoquées. Ces *Ballspring Physics*, au-delà de leur esthétique, intéressent particulièrement Cheng pour leur dimension sensorielle. Regrettant l’absence de contacts réels¹ entre les objets 3D dans les films d’animation ou les jeux vidéo, Cheng, convaincu que la vue du toucher est vectrice d’empathie, cherche à provoquer le choc des corps virtuels. Il se lance alors, avec son équipe, dans une série de simulations qui mettent à l’honneur les contacts physiques : de courtes capsules vidéos dans lesquelles de petits personnages virtuels se servent de leurs corps pour soulever des objets, se déplacer, ou interagir entre eux. En cela, la démarche de Cheng est très proche de celle de Diakur : créer de l’empathie pour des créatures virtuelles en les dotant d’un corps et, finalement, d’une certaine humanité. Mais l’artiste n’en reste pas là : poursuivant dans la voie des simulations sur *Unity*, Cheng s’éloigne légèrement de l’animation pour se rapprocher du jeu vidéo. Très inspiré par des jeux comme les *Sims*, il introduit peu à peu du sens, du conflit et des objectifs dans ses œuvres, poursuivant l’ambition de concevoir des simulations infinies, de « faire un jeu vidéo qui pourrait jouer à lui-même, de façon à atteindre cette vivante »². La création de sa trilogie *Emissaries* répond à ce projet de créer des écosystèmes virtuels, chacun des trois épisodes mettant en scène un même lieu virtuel constitué de nombreuses entités à travers différentes époques. La première œuvre, *Emissary in the Squat of Gods*, conçue en 2015 simule l’existence d’une tribu d’êtres humanoïdes vivant au pied d’un volcan menaçant d’entrer en éruption. La seconde œuvre, *Emissary Forks At Perfection*, conçue en 2015-2016, se déroule quelques milliers d’années après le premier *Emissary*, à l’intérieur du volcan qui est devenu un paradis fertile. Enfin, la dernière œuvre de la trilogie, *Emissary Sunsets The Self*, conçue en 2017, prend place des milliards d’années plus tard, alors que l’intelligence artificielle à son paroxysme a évolué en une substance qui s’est « mêlée avec le paysage pour former un atoll sentient »³. Pour ces simulations, Cheng conçoit tout un système des

¹ Dans le sens où les géométries qui se touchent n’ont pas d’effet les unes sur les autres, par opposition aux simulations physiques qui calculent la dynamique des chocs.

² Ma traduction de : « make a video game that could play itself so that it would achieve this aliveness », Ian Cheng, cité dans Henry Levinson, « A Great Anomic Era: IAN CHENG », *032c*, 19 septembre 2022, URL : <https://magazine.032c.com/magazine/a-great-anomic-era-ian-cheng>, consulté le 4 septembre 2026.

³ Ma traduction de : « and merged with the landscape to form a Sentient Atoll », Ian Cheng, « Emissaries », *iancheng.com*, à propos de *Emissary Sunsets the Self* (2017), URL : <https://iancheng.com/emissaries>, consulté le 4 septembre 2026.

règles et de comportements, administrant notamment l'existence sociale de la communauté quasi-humaine du premier *Emissary*. Bien que présente dans ces simulations, la trame narrative n'est pas centrale. Cheng compare ses œuvres aux caméras que les biologistes placent dans des espaces naturels avec l'espoir de capturer des images d'animaux sauvages. Dans ces films, la plupart du temps il ne se passe rien, mais au bout d'une dizaine heures un panda, ou n'importe quel autre animal, finit toujours par apparaître et réaliser un numéro spécial. C'est cet effet que Cheng cherche à reproduire : filmer un écosystème virtuel dans son existence souvent banale mais également capable de surprises. Dans cette optique, la trilogie *Emissaries* est exposée sur de grands écrans qui font office de fenêtres ouvrant sur des paysages virtuels. À la différence de Diakur qui crée des films visibles en ligne ou projetés dans des festivals, Cheng crée des installations physiques auxquelles il apporte un très grand soin. C'est que ces grands écrans sont le moyen pour l'artiste d'accomplir son objectif : « Make the viewer fall in love with a system ».

Les systèmes de Cheng reposant en grande partie sur les entités semi-autonomes qui les habitent, celui-ci apporte une grande attention à la création d'agents dans le sens que nous avons déjà évoqué : des entités artificielles capables de percevoir et d'agir dans leur environnement. En partant des travaux du neuroscientifique Marvin Minsky et de sa thèse selon laquelle l'esprit est une architecture d'agents indépendants, Cheng développe un modèle d'esprit pour les personnages de sa trilogie *Emissaries* composée de trois cerveaux en compétition : d'abord le « modèle de menace » qui réagit à la peur et au danger, ensuite le « modèle des besoins » qui gère les besoins primaires comme l'alimentation ou le repos, et enfin, le « modèle narratif » qui crée du sens. Cette architecture mentale correspond donc à une sorte de « parlement des agents » dans lequel chacun des trois modèles cherche à prendre le pouvoir sur les deux autres et, ainsi, contrôler l'action du personnage virtuel. Derrière cette architecture, l'ambition de Cheng n'est pas de créer une véritable intelligence mais plutôt un système de contrôle interne des entités peuplant ses simulations qui ne soit pas simplement aléatoire. Et ce travail sur la modélisation de l'esprit prend une nouvelle dimension dans une nouvelle œuvre, *BOB* pour *Bag Of Beliefs*, réalisée en 2018-2019 qui met en scène une créature étrange, à mi-chemin entre un serpent et une algue, évoluant dans un environnement virtuel. Après avoir travaillé sur des écosystèmes avec *Emissaries*, Cheng désire se concentrer sur le design d'une créature unique et plus complexe, simuler le métabolisme d'un organisme, la manière dont il vit, meurt, et interagit avec son environnement ; surtout, il cherche à créer une entité artificielle sentiente disposant d'une forme de persistance et d'une capacité évolutive. Cheng définit la sentience comme la capacité à être bouleversé et à régler ces bouleversements. Il décide donc de « centrer l'architecture cognitive de BOB sur la relation entre

les désirs et les croyances. Les croyances organisent les désirs. Les désirs agissent sur le monde. Le monde confirme ou bouleverse les croyances. »¹. Dans la continuité de ses précédents travaux, Cheng conçoit l'esprit de *BOB* sous la forme d'un agrégat d'agents : une dizaine de cerveaux en compétition, luttant pour leur capacité d'action.

BOB est doté d'une architecture d'intelligence artificielle composite singulière, composée d'un congrès de « démons » motivationnels et d'un moteur inductif neurosymbolique capable d'apprendre des croyances fondées sur des règles à partir d'expériences sensorielles. Chaque démon fonctionne comme une sous-personnalité qui tente, de façon obsessionnelle, de mener à bien sa propre micro-histoire. Un démon mangeur cherche de la nourriture, un démon de fuite évite les menaces perçues, un démon explorateur recherche ce qui est dépourvu de croyances, parmi bien d'autres. Ensemble, les démons se disputent le contrôle du corps de BOB.²

Chacun de ces cerveaux est doté d'un objectif et d'un narratif propres, porteurs de biais perceptifs. Ainsi, bien que reliés aux mêmes capteurs, les sous-agents qui composent l'esprit de *BOB* perçoivent différemment. Lorsque l'un des cerveaux l'emporte sur les autres, sa perception est transmise au « moteur d'inférence », l'organe central de l'esprit de *BOB* qui établit une croyance à partir de cette perception. Cette croyance prévaut et guide l'action de *BOB* jusqu'à ce qu'il se heurte à un état de fait en contradiction avec la croyance : face à ce bouleversement, *BOB* est alors forcé de modifier sa croyance. Les bonnes croyances permettent à *BOB* de survivre, d'échapper au danger ou de se nourrir efficacement. Mais celui-ci peut également développer de mauvaises croyances en commettant des erreurs d'inférences. Imaginons, par exemple, que *BOB* tombe sur une pomme rouge pour la première fois. Après une lutte interne, le cerveau de la nourriture s'impose et *BOB* mange la pomme. Il découvre alors qu'elle est comestible et qu'elle le nourrit convenablement. À la suite de cette expérience, il développe la croyance que les objets rouges sont comestibles. Peu après, *BOB* trouve une brique rouge sur son chemin. Sans hésiter, il croque dedans et se fait terriblement mal. Sa croyance sur les objets rouges change brutalement : ils sont dangereux et à éviter à tout prix. La prochaine fois que *BOB* croise une pomme il l'ignore, et la fois d'après encore alors même qu'il est affamé. *BOB* risque très vite de mourir de faim, à moins que sa croyance ne change à nouveau, par exemple à la découverte d'une pomme verte. Mais après avoir conçu

¹ Ma traduction de : « center BOB's cognitive architecture on the relationship between desires and beliefs. Beliefs organize desires. Desires act on the world. The world affirms or upsets beliefs », Ian Cheng, « BOB », *iancheng.com*, URL : <https://iancheng.com/BOB>, consulté le 4 septembre 2026.

² Ma traduction de : « BOB features a unique composite AI architecture composed of a congress of motivating "demons" and a neurosymbolic inductive engine capable of learning rule-based beliefs from sensory experiences. Each demon functions as a subpersonality who obsessively attempts to fulfill its own micro-story. An eater demon forages for food, a flight demon evades perceived threats, an explorer demon seeks that which lacks beliefs, among many others. Together the demons compete with one another for control of BOB's body », *Ibid*.

l'architecture cognitive de *BOB* d'une manière théorique, Cheng prend vite conscience qu'il doit doter sa créature d'un corps : l'agentivité de *BOB* repose d'abord sur son existence corporelle, clé de sa perception, nécessaire au développement de ses croyances, et de son action, nécessaire à l'exercice de ses désirs. Cheng rend compte de l'incarnation de *BOB* :

J'ai passé la première année à développer le corps de BOB pour qu'il croisse de manière procédurale, via un réseau de nœuds bille-ressort, vers une grande variété de morphologies. J'ai expérimenté différentes façons d'animer le corps de BOB, avant de concevoir un système locomoteur procédural combinant un comportement d'essaimage (*flocking*) de ses nœuds terminaux avec une redistribution constamment actualisée de son centre de gravité. J'ai développé un ensemble de capteurs physiques pour BOB : des capteurs externes, incluant des récepteurs de douleur sur chaque nœud, une détection de surétirement entre les nœuds, des capteurs visuels détectant le mouvement, la couleur, la forme, la texture, la composition élémentaire des sous-composants ; des capteurs internes, incluant l'énergie métabolique, l'intégrité constitutionnelle, ainsi que la capacité de l'estomac et de la vessie. Enfin, j'ai doté BOB d'un répertoire d'actions de base : se déplacer vers une cible, faire face, toucher, tenir, lâcher, pousser, frapper, lancer, mordre, suivre, exprimer des émotions faciales, tomber comme une poupée de chiffon (*ragdolling*), se figer, vomir, élaguer un segment corporel.¹

Basé sur le système des *Ballspring Physics* de *Unity* mentionné précédemment, le corps de *BOB* fait de lui un organisme fluide et composite : une sorte de serpent constitué de membres pouvant se combiner de manière procédurale. Le corps de *BOB* n'est pas figé, il peut se déformer, se diviser ou se recomposer : il évolue sans cesse. La mise au point de *BOB* constitue pour Cheng l'occasion de s'intéresser à l'intelligence artificielle incorporée qui présuppose un « sens de la signification orienté vers l'action »², « la capacité à percevoir le monde comme un ensemble d'opportunités d'action, et non comme un paysage d'objets platoniciens, objectifs »³. Pour *BOB*, c'est donc bien par l'existence corporelle qu'advient le sens, et, encore une fois, pour le spectateur, c'est du corps que surgit l'empathie. On ne se lasse pas des simulations physiques qui animent ce serpent élastique, de

¹ Ma traduction de : « I spent the first year developing BOB's body to procedurally grow via a network of ball-spring nodes into a wide variety of morphologies. I experimented with ways to animate BOB's body, eventually devising a procedural locomotive system that combined flocking behavior of its end nodes with a constantly updated redistribution of its center of gravity. I developed a set of physical sensors for BOB: external sensors including pain receptors on each node; over-stretch detection between nodes; vision sensors that detect movement, color, shape, texture, basic composition of subcomponents; internal sensors including metabolic energy, constitutional integrity, and stomach and bladder capacity. Finally, I gave BOB a repertoire of basic actions: moving to a target, facing, touching, holding, dropping, pushing, hitting, throwing, biting, tracking, facial emoting, ragdolling, freezing, vomiting, pruning a body segment. », *Ibid.*

² Ma traduction de : « action-oriented sense of meaning », Ian Cheng, cité dans Alex Zafiris, « A Coefficient of Luck: Ian Cheng in conversation with Alex Zafiris », *Maquette*, n° 3, Center for Collaborative Arts and Media, Yale University, URL : <https://yalemaquette.com/Q-A-Ian-Cheng>, consulté le 4 septembre 2026.

³ Ma traduction de : « the ability to perceive the world as opportunities for action, not as a landscape of platonic, objective objects », *Ibid.*

le regarder se mouvoir et agir dans son vivarium virtuel. Encore une fois, on éprouve la sensation privilégiée d'assister à l'émergence du vivant. Et bien que Cheng reconnaisse lui-même que son œuvre est encore loin de la sentience, les « émotions » de *BOB* ne laissent personne indifférent. Cheng raconte qu'à l'occasion d'une exposition, *BOB* a été si maltraité par les visiteurs via l'application mobile permettant d'interagir avec la créature, qu'il s'est rétracté sur lui-même, devenu tout petit, incapable de grandir à nouveau ; il a continué de vivre dans l'exposition sous cette forme atrophiée, incapable de faire quoi que ce soit, et tout le monde a eu un sentiment étrange de tristesse, voire de mauvaise conscience.

Avec ses réalisations artistiques, Cheng dit vouloir générer une « aliveness », qu'on pourrait traduire par « vivance », caractérisée par l'automatisme du mouvement interne et autonome. S'il utilise les simulations informatiques, c'est parce qu'elles permettent des comportements émergents — qui naissent du tout sans être contenus dans les parties —, à plus forte raison encore avec le *deep learning* et l'intelligence artificielle. Mais la simulation n'est qu'un moyen, la véritable pratique artistique de Cheng c'est le « worlding » qu'il théorise dans son essai *Emissary's guide to Worlding*¹. Celui-ci se définit comme « l'activité artistique d'un artiste individuel qui conçoit, incube, déclenche et nourrit un Monde vers l'*aliveness* »². Il s'agit non seulement de « étoffer le territoire du lieu, des personnages et de leur histoire, comme le ferait un romancier »³ mais aussi de créer « un lieu véritable que les gens peuvent habiter »⁴. Au début de son ouvrage, Cheng expose son ambition de « produire un Monde qui pourrait s'écrire lui-même, de la même manière que la nature s'écrit elle-même »⁵. Cheng s'intéresse particulièrement au paradoxe que surgit du *worlding* : comment faire coexister la finitude de l'auteur que présuppose toute création et l'infinité qui caractérise un monde ? C'est cette volonté de s'attaquer au concept paradoxal de « monde artificiel » qui motive Cheng :

Je voulais créer un Monde à partir de rien, comprendre l'artificialité propre à la fabrication d'un Monde, et devenir attentif au moment magique où un Monde commence à prendre une vie qui lui est

¹ Ian Cheng, *Emissaries Guide to Worlding*, Londres, Serpentine Galleries/Koenig Books, 2018

² Ma traduction de : « the artistic activity of an individual artist conceiving, incubating, triggering, and nurturing a World towards aliveness. », *Ibid.*, p. 21

³ Ma traduction de : « fleshing out the territory' of place, character, and backstory as a fiction writer would », Ian Cheng, cité dans Will Jennings, « Artist Ian Cheng explores the technological and aesthetic potential of AI », *Wallpaper**, 16 septembre 2022 (mis à jour le 28 septembre 2022), URL : <https://www.wallpaper.com/art/ian-cheng-life-after-bob-halle-am-berghain>, consulté le 4 septembre 2026.

⁴ Ma traduction de : « an actual place that people can inhabit », *Ibid.*

⁵ Ma traduction de : « produce a World that could 'write itself', the way nature writes itself », Ian Cheng, *Emissaries Guide to Worlding*, op. cit., p.14

propre, malgré son artificialité. Plus que tout, je voulais savoir comment le fait de fabriquer un Monde pourrait rendre les Mondes existants moins sacrés. Ensemble, en mon nom, les masques se sont mis à exercer une activité étrange et pourtant unifiée : le Worlding.¹

Dans cette optique, Cheng cherche d'abord à définir le concept de « monde » du point de vue du créateur, ce qui constitue une démarche originale, les mondes étant toujours déjà-là, toujours considérés du point de vue de leur habitants. Pour Cheng, un monde entre dans l'existence dès lors qu'un créateur commence à y croire et, immédiatement, se surprend à « imaginer un autre plaisir : abandonner le rôle de créateur et être une personne vivant dans la croyance, bénéficiaire de son potentiel, un croyant »². La création de monde implique dès lors, à son origine, le dépassement de son créateur. À partir de là, Cheng propose une définition simple : « Un Monde est un future dans lequel on peut croire: qui promet de survivre à son créateur et de continuer à générer du *drama* »³. Cette définition implique deux caractéristiques fondamentales d'un monde : d'une part, l'autonomie — le dépassement et la survivance au créateur — et, d'autre part, l'*aliveness*, ou la « vivance » — la capacité à générer de la tension dramatique, des frictions, à entretenir la métastabilité qui caractérise la vie. Les mondes qui ne survivent pas à leurs créateurs ou ne sont plus en mesure de produire de la métastabilité s'éteignent d'une mort lente. Le *worlding* c'est donc le processus de conception d'une croyance qui devient un monde, qui développe une *vivance* et une autonomie. Cheng s'inspire du travail de James Carse, chercheur en histoire des religion, et de son livre *Finite and infinite games*⁴, pour développer sa théorie du worlding. Carse fait la différence entre « jeux finis » et « jeux infinis » : les premiers ont un but, gagner, tandis que les seconds n'ont d'autres objectif que la poursuite infinie du jeu. Le paradigme du jeu infini, c'est le jeu de l'évolution, la nature qui constitue l'unique forme de jeu infini selon Carse. Pour Cheng, en revanche, le concept de « jeu infini » est utile pour désigner l'ensemble des mondes construits socialement dont nous nous servons pour organiser la réalité et y donner du sens :

Un Monde est une construction. Il n'est rien comparé au véritable jeu infini qu'est la Nature, mais il est infini *suffisamment* — car il soutient les qualités d'un jeu infini assez longtemps et de façon assez

¹ Ma traduction : « I wanted to create a World from nothing, to understand the artificiality of making a World, and to become attentive to the magical moment when a World begins to take on a life of its own, despite its artificiality. Most of all, I wanted to know how making a World might make existing Worlds seem less sacred. Together on my behalf, the masks began exercising a strange yet unified activity: Worlding. », *Ibid.*

² *Ibid.*

³ Ma traduction de : « A World is a future you can believe in: One that promises to survive its creator, and continue generating drama. », *Ibid.*, p.17

⁴ James P. Carse, *Finite and Infinite Games: A Vision of Life as Play and Possibility*, first Free Press paperback edition (Free Press, 2012).

surprenante pour que les humains le traitent avec le statut d'être vivant. Un Monde est une chose vivante artificielle, mais une chose vivante tout de même.¹

Les mondes sociaux, à l'instar des religions, naissent de la croyance et de l'entretien de la foule des individus qui y prennent part sur des générations. Il y a bien sûr des mondes dont l'origine est imputable à une seule personne, par exemple les fictions, mais ceux-ci ne deviennent véritablement mondes qu'au moment où, se détachant de leur auteur, ils sont investis par d'autres et atteignent la complexité nécessaire pour être en mesure de générer de la « ongoingness », une inertie qui va de l'avant. Mais dès lors que les mondes sont toujours la création de la multitude humaine qui les habite, comment le *worlding* pourrait-il être un art ? Comment pourrait-il être l'œuvre d'un créateur unique, l'artiste ? Si Cheng reconnaît l'impossibilité, encore récente, de concevoir un monde pour une personne seule, dans l'espace d'une seule vie, c'était sans compter sur l'émergence des nouvelles technologies de simulation et d'intelligence artificielle. Soudain, les agents autonomes apparaissent comme une alternative aux humains pour habiter les espaces virtuels, les remplir d'un semblant d'intersubjectivité et les féconder pour en faire des mondes. L'IA ouvre alors la porte à la création de mondes par un seul artiste, rendant possible le *worlding* comme genre artistique : « Il y a ce sentiment que créer un Monde — le *Worlding* — pourrait bien être à la portée d'un artiste seul »².

Au début du processus de *worlding*, le monde est fragile, purement artificiel et dépendant de son créateur. Mais peu à peu, il prend vie et développe une autonomie, gagne en *aliveness* et attire d'autres habitants. Cheng propose une équation pour mesurer le degré de *vivance* d'un monde : son degré d'utilisation (« evidence usage ») divisé par l'emprise du créateur (« creator's hold »). Plus l'usage du monde dans toutes ses manifestations est répandu, c'est-à-dire, plus il y a des gens qui l'incarnent et le font vivre, et moins l'emprise du créateur est forte, plus le monde est en vie. À l'inverse, moins le monde est habité et plus le contrôle qu'y exerce son créateur est important, et plus le monde ressemble à un zombie. Un monde atteint l'autonomie totale lorsque l'emprise du créateur tombe à zéro, et il meurt lorsqu'elle monte à un. Pour Cheng, le *worlding* se divise donc en deux grandes étapes : d'abord insuffler la vie (*aliveness*) et puis accorder l'autonomie : « Le *Worlding* est d'abord l'acte de créer une vie, puis, dans un second temps, de laisser cette vie se vivre

¹ Ma traduction de : « A World is a construction. It is nothing compared to the true infinite game of Nature, but it is infinite enough because it sustains the qualities of an infinite game long enough and surprising enough for humans to treat it with the status of being alive. A World is an artificial living thing, but a living thing nonetheless », Ian Cheng, *Emissaries Guide to Worlding*, op. cit., p. 19

² Ma traduction de : « There is the feeling that creating a World – Worlding – might be just within reach of an individual artist. », *Ibid.*, p. 20

elle-même. La première partie concerne l'atteinte de l'*Aliveness*. La seconde concerne l'octroi de l'*Autonomie*. »¹. Ce projet de *worlding* est particulièrement évident dans le moyen métrage d'animation d'une cinquantaine de minutes *Life After BOB* conçu par Cheng comme une suite à son œuvre *BOB*. Entièrement réalisé dans *Unity*, *Life After BOB* constitue une petite prouesse technique : c'est le premier film de cette longueur à avoir été entièrement produit dans le moteur. Si Cheng a choisi de réaliser ce film dans un logiciel d'abord conçu pour le jeu vidéo, c'est qu'il tenait à inscrire ce projet dans la continuité de son travail précédent sur les simulations et de l'inscrire dans sa démarche du *worlding*. Et, de fait, Cheng a pu profiter des avantages du realtime et des moteurs de jeu, déjà évoqués, pour rapprocher l'expérience de sa réalisation de l'interactivité du tournage ou du jeu vidéo :

J'ai appris qu'il est désormais possible de traiter le cinéma comme on traiterait la fabrication de logiciels. Plutôt que de penser les médias narratifs comme une super-œuvre parfaite, coûteuse et laborieuse, il devient possible d'itérer rapidement et à moindre coût, et de développer de nouvelles fonctionnalités para-narratives qui élargissent l'expérience du spectateur s'il devient fan et va les chercher.²

D'ailleurs, la technique indiquée dans le descriptif du film est celle de « Realtime Live Animation » : le film n'est plus un fichier vidéo pré-calculé, mais un programme, plus proche en cela d'une représentation théâtrale. Et, en effet, *Life After BOB* n'est pas présenté comme un film classique, il est exposé avec une scénographie particulière. À Berlin, il est montré dans la célèbre boîte de nuit du *Berghain*, reconvertie pour l'occasion en lieu d'exposition. Avant d'accéder à l'espace de projection, les visiteurs vivent une expérience de son et lumières qui fait office de conditionnement physique. Et puis, après le visionnage des 48' d'animation sur grand écran, les spectateurs rejoignent un deuxième espace où ils peuvent revoir le film, cette fois en interagissant avec lui. Tirant parti de sa fabrication dans un moteur de jeu, Cheng a conçu une application permettant aux spectateurs de mettre en pause n'importe quel moment du film et de prendre la main sur la réalisation. Le téléphone fait office de télécommande : il est possible de cliquer sur des éléments de l'image, de contrôler la caméra pour changer l'angle de vue ou le focus, ou encore d'interagir avec le *wiki* de l'œuvre pour rédiger la notice biographique de certains des personnages... Ici, le film

¹ Ma traduction de : « Worlding is firstly the act of creating a life, then secondly letting that life live itself. The first part is about achieving Aliveness. The second part is about granting Autonomy. », *Ibid.*, p. 23

² Ma traduction de : « I learned that it is now possible to treat filmmaking like software-making. Instead of thinking of narrative media as a perfected, expensive, laborious super-artwork, it can now be about iterating quickly and inexpensively, and developing new para-narrative features that expand the viewer's experience if they become fans and seek it. », Ian Cheng, cité dans Ann Marie Alanes, « Can AI Unlock the Unconscious? », *Right Click Save*, 14 octobre 2022, URL : <https://www.rightclicksave.com/article/can-ai-unlock-the-unconscious>, consulté le 4 septembre 2026.

échappe à son créateur : ce n'est plus une œuvre figée, mais au contraire, il se déploie comme un monde habitable par les spectateurs qui deviennent de nouveaux auteurs. Cheng nomme cette façon de regarder des films « worldwatching » poursuivant un vieux rêve : « laisser le spectateur mettre en pause un épisode et commencer à improviser une conversation avec un personnage ou un narrateur. »¹. Ce n'est plus tout à fait un film qu'on a devant les yeux, mais un monde.

Finalement, la pratique du *worlding* permet d'apprendre à aimer le changement permanent, à accueillir le chaos du monde. Cheng appelle à un « art qui est vivant et dynamique »² pour nous aider à mieux appréhender et représenter le mouvement constant des choses. Un enjeu de taille se dessine alors pour les nouvelles technologies. Les simulations informatiques, en particulier, pourraient constituer de nouveaux grands récits, des « jeux infinis » alternatifs à ceux déjà en marche comme les religions. Surtout, elles inaugurent un moment nouveau dans l'histoire humaine où n'importe quel individu, désormais doté du pouvoir de démiurge, n'est plus seulement condamné à être *jeté-au-monde-déjà-là*, mais est en mesure de faire advenir de nouveaux mondes. Poursuivant l'ambition la plus ancienne de l'art, le *worlding* est une manière d'élargir la condition humaine.

¹ Ma traduction de : « let the viewer pause an episode and begin to improvise a conversation with a character or narrator. », *Ibid.*

² Ma traduction de : « art that is alive and dynamic », *Ibid.*

Bibliographie

AIKEN Howard Hathaway, « A Proposed Automatic Calculating Machine (1937) » in *Ideas That Created the Future: Classic Papers of Computer Science*, Harry R. Lewis (éd.), s.l., The MIT Press, 2021.

AKENINE-MÖLLER Tomas, HAINES Eric, HOFFMAN Naty, PESCE Angelo, IWANICKI Michał, HILLAIRE Sébastien et HOFFMAN Nathaniel, *Real-time rendering*, Fourth edition., Boca Raton London New York, CRC Press, Taylor & Francis Group, « An A K Peters book », 2018, 1178 p.

ALANES Ann Marie, « Can AI Unlock the Unconscious? », in *Right Click Save*, 14 octobre 2022, en ligne : <https://www.rightclicksave.com/article/can-ai-unlock-the-unconscious>.

ALBERA François, « Notes sur les rapports de l'Office catholique international du cinéma et la filmologie », in *Cinémas*, n° 2-3, vol. 19, 2009.

ARISTOTE, *Métaphysique*, Ed. minor., Paris, J. Vrin, 1991.

BACHELARD Gaston, *La formation de l'esprit scientifique*, 16. éd., Paris, Vrin, « Bibliothèques des textes philosophiques », 1999, 256 p.

BAKER Lynne Rudder, *The metaphysics of everyday life: an essay in practical realism*, Cambridge, Cambridge university press, « Cambridge studies in philosophy », 2007.

BALÁZS Béla, *Béla Balázs: early film theory: Visible man and the spirit of film*, traduit par Rodney Livingstone, New York, Berghahn Books, 2011.

BARBARAS Renaud, *La perception. Essai sur le sensible*, Hatier., Paris, s.n., 1994.

BARNIER Martin et JULIER Laurent, *Une brève histoire du cinéma : 1895-2020*, S.l., Fayard/Pluriel, 2021.

BARTHES Roland, *La chambre claire: note sur la photographie*, Nachdruck., Paris, Gallimard Seuil, « Cahiers du cinéma », 2016, 192 p.

- BAUER Friedrich L., *The Computer - My Life*, Berlin, Heidelberg, Springer Berlin / Heidelberg, 1993.
- BAZIN André, *Qu'est-ce que le cinéma ?*, Paris, Cerf, « 7e art », 1985.
- BERGSON Henri, *L'évolution créatrice*, Réimpression de la 12^e éd., Paris, Presses universitaires de France, « Quadrige », 2013.
- BOYER Elsa, *Le conflit des perceptions*, Paris, Éditions MF, « Collection Inventions », 2015, 203 p.
- BRESSON Robert, *Notes sur le cinématographe*, Paris, Gallimard, « Collection Folio Texte intégral », 1995, 137 p.
- BUCCHIONI Guillaume et DECLOS Alexandre, *The Constitution of Virtual Objects*, Collège de France, s.n., 2023.
- CANGELOSI Angelo, *Simulating the Evolution of Language*, 1st ed., London, Praxis, 2002, 1 p.
- CANI Marie-Paule, « Mondes virtuels : genèse des nouveaux mondes » in *La Création*, s.l., Odile Jacob, « Le Temps des savoirs », 2005, p. 97-110.
- CERUZZI Paul E., *Computing: a concise history*, Cambridge (Mass.), MIT press, « The MIT press essential knowledge series », 2012.
- CHALMERS David, *Reality+: virtual worlds and the problems of philosophy*, First edition., New York, NY, W. W. Norton & Company, 2022.
- _____, « The Virtual as the Digital », in *Disputatio*, vol. 11, 1 décembre 2019, p. 453-486.
- _____, « The Virtual and the Real », in *Disputatio*, vol. 9, mars 2018, p. 309-352.
- CHENG Ian, « BOB », *iancheng.com*, en ligne : <https://iancheng.com/BOB>.
- _____, « Emissaries », *iancheng.com*, en ligne : <https://iancheng.com/emissaries>.
- _____, *Emissaries Guide to Worlding: How to Navigate the Unnatural Art of Creating an Infinite Game by Choosing a Present, Storytelling its Past, Simulating its Futures, and Nurturing its Changes*, éd. Rebecca Lewin, Joseph Constable et Veronica So, Londres, Serpentine Galleries / Fondazione Sandretto Re Rebaudengo / Koenig Books, 2018.

____, *Ian Cheng DMA Lecture Series 2017-18*, conférence donnée à UCLA Design Media Arts, 7 novembre 2017, chaîne YouTube UCLA Design Media Arts, en ligne : <https://youtu.be/gBOfkWTUnvI>.

COUCHOT Edmond, *Automates, robots et humains virtuels dans les arts vivants*, Saint-Denis, Presses universitaires de Vincennes, « Théâtres du monde », 2022, 220 p.

DELEUZE Gilles, *Cinéma. 1, L'image-mouvement*, Paris, Éditions de Minuit, 2013.

DESCARTES René, *Discours de la méthode*, Paris, Flammarion, « GF », 2000, 189 p.

DESCOLA Philippe, *Les formes du visible: une anthropologie de la figuration*, Paris, Éditions Points, « Points Essais », 2023.

DEWEY John, *L'art comme expérience*, Paris, Gallimard, « Folio », 2010.

____, *Art as experience*, Perigee trade paperback edition., New York, Berkley Publishing Group, 2005.

DIAKUR Nikita, dossier de presse de *Backflip* [PDF en ligne], 2022, en ligne : https://backflip.training/press/backflip_info.pdf.

DUPOND Pascal, *Dictionnaire Merleau-Ponty*, Paris, Ellipses, « Dictionnaire », 2007.

FENG K. J., McDONALD David W. et ZHANG Amy X., « Levels of Autonomy for AI Agents », in *arXiv preprint arXiv:2506.12469*, 2025.

FLORIDI Luciano, *Information: a very short introduction*, Oxford, Oxford University press, « Very short introductions », 2010.

FUNG Pascale, BACHRACH Yoram, CELIKYILMAZ Asli, CHAUDHURI Kamalika, CHEN Delong, CHUNG Willy, DUPOUX Emmanuel, JÉGOU Hervé, LAZARIC Alessandro et MAJUMDAR Arjun, « Embodied AI Agents: Modeling the World », in *arXiv preprint arXiv:2506.22355*, 2025.

GAUDIN Antoine, *L'espace cinématographique: Esthétique et dramaturgie*, Paris, Armand Colin, 2022.

- GOMBRICH Ernst Hans, *Histoire de l'art*, Nouv. éd., Paris, Flammarion, 1994.
- GUPTA Agrim, SAVARESE Silvio, GANGULI Surya et FEI-FEI Li, « Embodied intelligence via learning and evolution », in *Nature Communications*, n° 1, vol. 12, 6 octobre 2021, p. 5721.
- HA David et SCHMIDHUBER Jürgen, « Recurrent World Models Facilitate Policy Evolution », s.l., Curran Associates, Inc., 2018, vol. 31/.
- HERZOG Werner, *Conquête de l'inutile*, Nantes [Paris], Capricci Centre Pompidou, « Collection dirigée par Emmanuel Burdeau », 2009.
- HITCHCOCK Alfred et TRUFFAUT François, *Le cinéma selon Alfred Hitchcock*, Paris, Robert Laffont, 1966.
- HOGG Trevor, « Focusing on virtual cinematography », in *VFXV*, 2023 p.
- KANERVISTO Anssi, BIGNELL Dave, WEN Linda Yilin, GRAYSON Martin, GEORGESCU Raluca, VALCARCEL MACUA Sergio, TAN Shan Zheng, RASHID Tabish, PEARCE Tim, CAO Yuhan, LEMKHENTER Abdelhak, JIANG Chentian, COSTELLO Gavin, GUPTA Gunshi, TOT Marko, ISHIDA Shu, GUPTA Tarun, ARORA Udit, WHITE Ryen W., DEVLIN Sam, MORRISON Cecily et HOFMANN Katja, « World and Human Action Models towards gameplay ideation », in *Nature*, n° 8051, vol. 638, 1 février 2025, p. 656-663.
- KIELA Douwe, BULAT Luana, VERO Anita L. et CLARK Stephen, « Virtual embodiment: A scalable long-term strategy for artificial intelligence research », in *arXiv preprint arXiv:1610.07432*, 2016.
- LATOUR Bruno, *Nous n'avons jamais été modernes: essai d'anthropologie symétrique*, Nachdr., Paris, Editions La Découverte [u.a.], 2010, 206 p.
- LECUN Yann et COURANT, « A Path Towards Autonomous Machine Intelligence Version 0.9.2, 2022-06-27 », s.l., s.n., 2022.
- MALABOU Catherine, *Métamorphoses de l'intelligence: du QI à l'IA*, Paris, PUF, « Quadrige », 2021.
- MALRAUX André, *Le musée imaginaire*, Paris, Gallimard, « Collection folio Essais », 2006, 285 p.

- ____, « Esquisse d'une psychologie du cinéma », in *Cinématographe*, n° 24, 1977.
- MANOVICH Lev, *The language of new media*, Cambridge, Mass., MIT Press, 2001.
- MANOVICH Lev et ARIELLI Emanuele, *Artificial Aesthetics*, s.l., s.n., 2024.
- MERLEAU-PONTY Maurice, *Le visible et l'invisible: suivi de Notes de travail*, Paris, Gallimard, « Tel », 2016, 359 p.
- ____, *Phénoménologie de la perception*, s.l., Editions Gallimard, 2013, 637 p.
- ____, *Sens et non-sens*, Nachdr., Paris, Gallimard, « Bibliothèque de philosophie », 1996, 225 p.
- ____, *L'œil et l'esprit*, Paris, Gallimard, « Collection folio Essais », 1989, 92 p.
- PANOFSKY Erwin, BALLANGÉ Guy et DALAI EMILIANI Marisa, *La perspective comme forme symbolique: et autres essais*, Paris, Les Éditions de Minuit, 1975, 273 p.
- PASCAL Blaise, *Pensées*, Paris, Flammarion, « GF », 2015.
- Peng, Xue Bin, Pieter Abbeel, Sergey Levine, et Michiel van de Panne. « DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills ». *ACM Transactions on Graphics* 37, n° 4 (2018): 1-14. <https://doi.org/10.1145/3197517.3201311>.
- QUÉAU Philippe, *Le virtuel: vertus et vertiges*, Paris, Champ Vallon ; Institut national de l'audiovisuel, « Collection Milieux », 1993, 215 p.
- SERRA Albert, « Par les temps qui courent ».
- SIMONDON Gilbert, *Du mode d'existence des objets techniques*, Paris, Aubier, 2012.
- SOHL-DICKSTEIN Jascha, WEISS Eric, MAHESWARANATHAN Niru et GANGULI Surya, « Deep unsupervised learning using nonequilibrium thermodynamics », s.l., pmlr, 2015.
- SOKOLOWSKI Robert, *Introduction to phenomenology*, Cambridge (GB) New York Melbourne, Cambridge university press, 2000.

- STEELS Luc, « Grounding Symbols through Evolutionary Language Games » in *Simulating the Evolution of Language*, Angelo Cangelosi et Domenico Parisi (éd.), London, Springer London, 2002, p. 211-226.
- TRICLOT Mathieu, *Philosophie des jeux vidéo*, Paris, la Découverte, « La Découverte-poche », 2017.
- TURING A. M., « On Computable Numbers, with an Application to the Entscheidungsproblem », in *Proceedings of the London Mathematical Society*, n° 1, vol. s2-42, 1937, p. 230-265.
- VALEVSKI Dani, LEVIATHAN Yaniv, ARAR Moab et FRUCHTER Shlomi, « Diffusion models are real-time game engines », in *arXiv preprint arXiv:2408.14837*, 2024.
- VAN DAMME Charles, « Le numérique, le virtuel et l'effet cinéma », in .
- VIAL Stéphane, *L'être et l'écran. Comment le numérique change la perception*, Paris, Presses Universitaires de France, 2013, 336 p.
- VON NEUMANN John, *The computer & the brain*, New Haven, Conn., Yale University Press, 2012.
- WIENER Norbert, *Cybernetics: or, Control and communication in the animal and the machine*, Second edition, 2019 reissue., Cambridge, Massachusetts, The MIT Press, 2019, 303 p.
- WITTGENSTEIN Ludwig, *Recherches philosophiques*, Paris, Gallimard, 2014.
- WOOLDRIDGE Michael J., *A brief history of artificial intelligence: what it is, where we are, and where we are going*, First U.S. edition., New York, Flatiron Books, 2021, 262 p.