Chapitre 1

La vision et la perception humaines

Chapitre rédigé par Francis SCHMITT

1.1 Anatomie du système visuel humain

1.1.1 L'œil

L'œil est le récepteur de la vision [LeGrand, 1956, Pirenne et Crouzy, 1972]. Il reçoit les rayons lumineux, en forme une image et transmet celle-ci au cerveau par le nerf optique, cf. figure 1.1.

L'image est formée à partir du cristallin qui joue le rôle d'une lentille de focale variable, sur la rétine, couche photosensible qui tapisse le fond de l'œil (surface à peu près sphérique). L'iris contrôle le diamètre de la pupille, comme un diaphragme automatique contrôlant l'énergie arrivant sur la rétine. Les muscles oculomoteurs assurent l'orientation de l'œil et la convergence des deux yeux sur le point d'intérêt. Convergence et focalisation sont généralement couplées dans la vision réflexe.

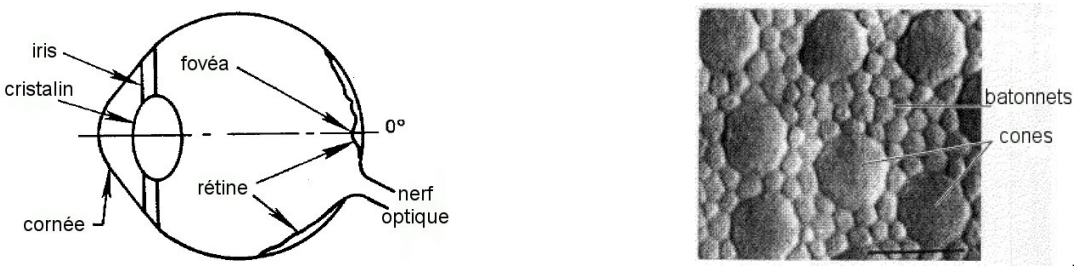


FIG. 1.1 – Schéma de l'œil humain et photographie du fond de la rétine.

Le champ visuel n'est pas circulaire, il est de 90° du côté temporal, de 60° du côté nasal, de 50° et 80° vers le haut et vers le bas.

La rétine est le transducteur qui transforme le signal optique en signaux électriques. La surface de la rétine est très hétérogène, aussi bien en nature de récepteurs qu'en densité, cf. figure 1.1. Cette hétérogénéité permet à des zones différentes d'assumer des fonctions complémentaires.

Le centre de la rétine assure la vision colorée et la vision des détails à fort niveau d'éclairement (vision photopique). Le pourtour assure la vision achromatique à faible niveau de lumière (vision scotopique), la vision des

mouvements rapides, la perception globale des scènes et la vision achromatique à faible niveau de lumière (vision scotopique). L'acuité visuelle est maximale sur la fovéa (zone de la rétine correspondant à l'axe du regard). La zone d'où part le nerf optique ne capte pas la lumière (elle est appelée tache aveugle).

La rétine est constituée de nombreux types de cellules nerveuses : les cellules photo-réceptrices (cônes et batonnets), les cellules bipolaires, ganglionnaires et amacrines. Ces cellules constituent 2 voies croisées : la voie directe et la voie indirecte, interconnectées de façon complexe. Dans la fovéa, la rétine adopte un schéma beaucoup plus simple : il n'y a que des cônes (et pas de batonnets) et chaque cône n'est connecté qu'à une cellule bipolaire elle même connectée à une seule cellule ganglionnaire.

L'acuité visuelle baisse régulièrement lorsqu'on s'éloigne de la fovéa.

Les cônes sont principalement concentrés dans la fovéa et leur densité décroît très rapidement vers le pourtour de l'œil. Ils sont au nombre de 6 à 7 millions. Dans la fovéa ils sont très petits (de 1 à 2 microns) et très serrés (entre 2 et 10 microns). Les cônes sont peu sensibles, mais une même fibre nerveuse ne regroupe qu'un tout petit nombre de cônes, assurant une bonne discrimination angulaire. Dans toutes les opérations de traitement d'image, on considère que l'image est observée en vision par les cônes. La vision chromatique est assurée par 3 types de cônes dont la sensibilité diffère selon le type de pigment qu'ils contiennent. Ces cônes sont appelés L, M et S en fonction des longueurs d'ondes auxquels ils ont leur pic de sensibilité : L= *Long* autour du jaune vert (585 nm), M= *Middle* autour du vert (550 nm) et S=*Short* autour du bleu (430 nm).

Les bâtonnets sont au contraire plus densément localisés sur une couronne centrée à 20° de la fovéa environ. Leur densité est nulle à la fovéa mais demeure très forte encore à plus de 60° de celle-ci. Ils sont très nombreux (de 80 à 150 millions), d'un diamètre moyen de 3 microns. Ils ont une très bonne réponse temporelle (c'est en périphérie de l'œil que passent les étoiles filantes), une très bonne sensibilité (ils assurent la vision nocturne - scotopique). Ils ont tous la même réponse spectrale (à peu près indépendante de la longueur d'onde) (« la nuit tous les chats sont gris. »).

1.1.2 Les voies optiques

Elles sont constituées de l'ensemble des cellules nerveuses qui partent de la rétine et aboutissent au cortex visuel primaire.

Le nerf optique tout d'abord part de l'œil et aboutit au chiasma optique, une structure où se croise une partie et se mélange une autre partie des fibres des nerfs optiques issus de chacun des yeux.

Sorties du chiasma optique les fibres se distribuent soit vers le Tubercule Quadrijumeau Antérieur soit vers le Corps Genouillé Latéral puis ensuite vers les aires du Cortex Visuel Primaire situées à l'arrière des lobes occipitaux.

(... à compléter ...)

1.1.3 Les aires supérieures de la vision

(... à compléter ...)

1.2 Comment percevons nous les images ?

Ce n'est pas en général le genre de question que l'on se pose lorsque l'on est face à un écran de télévision ou de cinéma ou lorsque l'on est plongé dans des bandes dessinées. Le système visuel fournit déjà à notre cerveau une quantité d'information suffisamment grande à traiter pour qu'on ne se pose pas encore le problème supplémentaire de savoir comment tout cela fonctionne. Ça marche et ça marche même bien ; merci dame Nature ! Ce type de préoccupation est en fait laissé au soin de diverses espèces de coupeur de cheveux en quatre : psychologues, physiologistes et autres individus comme ces ingénieurs qui traitent des images.

Sachant que l'image est un des supports physiques les plus privilégiés pour transmettre un message à notre cerveau, interrogeons-nous donc sur la manière dont ce message est transmis, décodé et synthétisé par le système nerveux. C'est une question extrêmement complexe qui n'est encore que très partiellement élucidée et qu'on va essayer de débroussailler grossièrement. Les divers types d'approche du phénomène de vision sont comparés dans [Gordon, 1998]. Il n'est pas question de faire ici un large tour d'horizon de l'état de nos connaissances sur le comportement du système visuel face à l'information contenue dans les images. Tout au plus allons-nous essayer de donner une vague vue synthétique qui n'évitera naturellement pas l'écueil d'une schématisation excessive. Le but, en fait, est de sensibiliser le lecteur sur le phénomène de la perception visuelle. Nous espérons que nous aurons ainsi l'occasion de lui procurer quelques sujets de distraction et de réflexion. Il pourra également se référer à [Gregory, 1966], [Cornsweet, 1973], [Murch, 1973] ou [Carterette et Friedman, 1975].

On peut distinguer dans le traitement de l'information « image » par notre système visuel deux procédures distinctes : une qui correspondrait à un comportement passif du système et l'autre à un comportement actif. Nous allons voir successivement à quoi correspondent ces deux aspects, sans cependant insister trop sur le premier qui est un tantinet technique et qui pourrait par là paraître un peu trop rébarbatif.

1.2.1 Comportement passif du système visuel

C'est ce comportement qu'étudient essentiellement les physiologistes et les ingénieurs qui mettent au point de nouveaux systèmes de traitement d'image. Ces derniers personnages ont besoin de bien connaître les propriétés statiques et dynamiques du système visuel pour pouvoir adapter, en fonction de ces contraintes, les caractéristiques physiques de l'image comme sa dimension, sa résolution, sa luminosité, son contraste... L'étude de ces propriétés se fait par des expériences de psychophysique : un individu placé dans des conditions d'observation parfaitement spécifiques regarde un signal donné, appelé stimulus, pouvant prendre des formes très diverses suivant la nature du phénomène étudié. Le problème est de relier les grandeurs physiques de ce stimulus aux grandeurs perceptives de la réponse qu'il engendre chez l'observateur. Dans la pratique on distingue deux types d'expériences : celles de détection et celles de perception. La réponse du système visuel est en effet différente suivant que l'on met en jeu de petites ou de grandes variations du signal. Lorsque celles-ci sont faibles, l'important est de savoir si elles sont au-dessus ou en-dessous du seuil de détection du système visuel. L'observateur aura par conséquent vu ou non la variation du signal et pourra donner une réponse assez objective sans trop se créer d'angoisses métaphysiques. Dans les expériences de perception où les variations du signal sont très grandes, le problème est plus délicat.

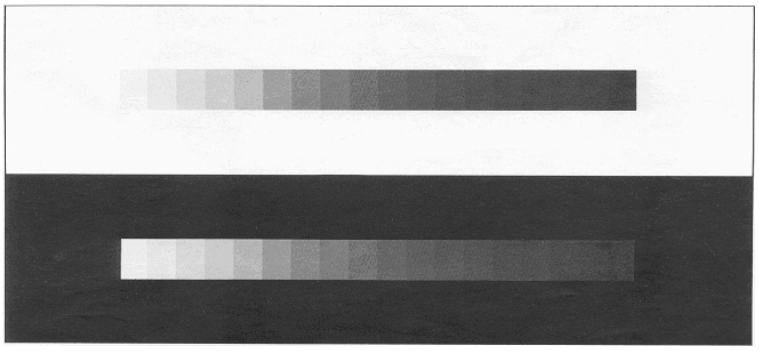


FIG. 1.2 – La perception des différentes gradations d'une échelle de gris dépend du niveau lumineux du fond.

Illustrons cela par un exemple et demandez-vous comment vous procéderiez pour fabriquer une échelle de gris d'une dizaine de tons allant du blanc au noir et dont les écarts de luminosité vous paraîtraient le plus homogène possible, d'un point de vue perceptif bien sûr. Vous vous apercevriez alors que cet exercice vous demande des processus intellectuels beaucoup plus élaborés que celui de fournir une réponse à la simple question « vu ou pas vu ? ». Il

vous faudra entre autres définir un critère de jugement qui vous permettra après entraînement de mieux sélectionner vos gris. Celà souligne l'aspect plus subjectif de telles expériences de perception comparées à celles permettant de mesurer les seuils de détection. La figure 1.2 montre l'importance des conditions selon lesquelles on opère et qu'il est nécessaire de bien contrôler : en changeant la luminosité du fond on modifie sérieusement la sensation que procure une échelle de gris déterminée. Le comportement passif du système visuel s'explique globalement par le fait que c'est essentiellement un système biophysique. Ses propriétés découlent des caractéristiques de sa structure, à commencer par l'œil et la rétine, pour finir avec l'immense complexité de ses connexions nerveuses (cf. section 1.1.2). Ce comportement peut donner lieu à quelques phénomènes particulièrement démonstratifs comme ceux des figures 1.2 à 1.5, dûs aux interactions spatiales entre les régions de l'image ayant des luminosités différentes. Ces exemples indiquent à quel point l'information « image » une fois prise en charge par le système visuel peut être rapidement transformée et déformée. indexcontraste simultané

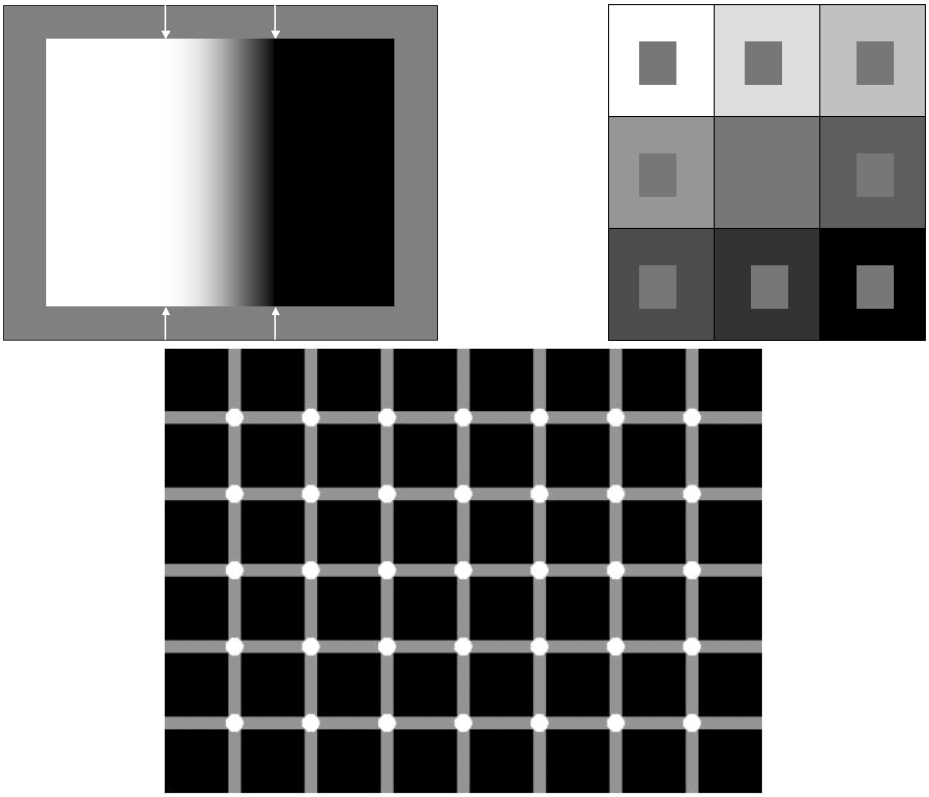


FIG. 1.3 – Phénomènes dus aux interactions spatiales dans l'image.
En haut à gauche une bande de Mach (c'est à dire un renforcement local du contraste) apparaît à la fin d'une variation spatiale continue de la luminance et est plus claire que le fond blanc sur lequel elle se situe.
En haut à droite, le contraste simultané est un exemple classique de l'interaction de la couleur d'un fond sur celle d'un médaillon : les neufs petits carrés ont le même niveau de gris.
En bas, le centre des croisements de la grille blanche d'Hermann paraît plus sombre excepté celui que l'on fixe du regard : la partie centrale de la rétine ne traite pas l'image de la même manière que sa partie périphérique.

A la figure 1.2 les plages dans l'échelle de gris ne sont pas perçues comme uniformes alors qu'elles le sont physiquement.Les trois phénomènes de la figure 1.3, trouvent leur origine aux niveaux des interconnexions nerveuses des cellules rétiniennes qui effectuent un véritable filtrage linéaire. Mais dès qu'intervient un contour, des phénomènes non linéaires (figure 1.4) peuvent se produire. Les contours dans certains cas peuvent procurer des

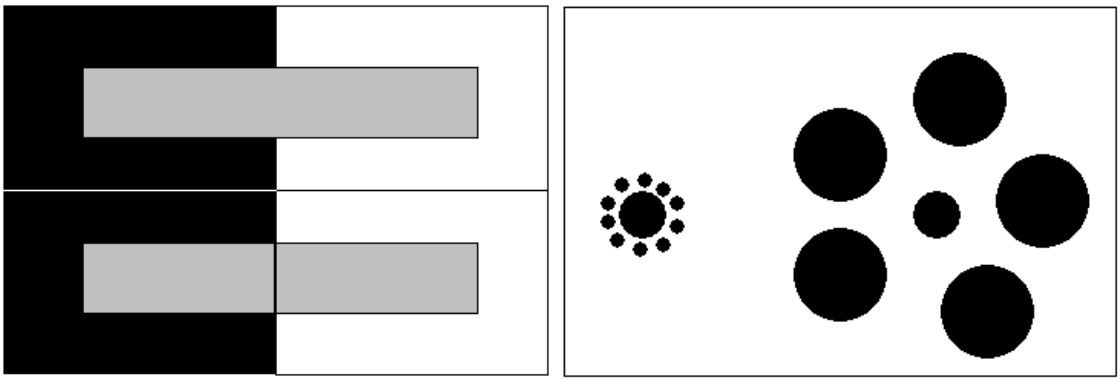


FIG. 1.4 – Phénomènes dus aux interactions spatiales dans l'image. À gauche, la seule adjonction d'un contour au milieu de la barre grise au motif du haut permet d'obtenir un net contraste simultané (motif du bas). À droite, dans l'illusion de Titchener les deux cercles centraux sont de taille égale.

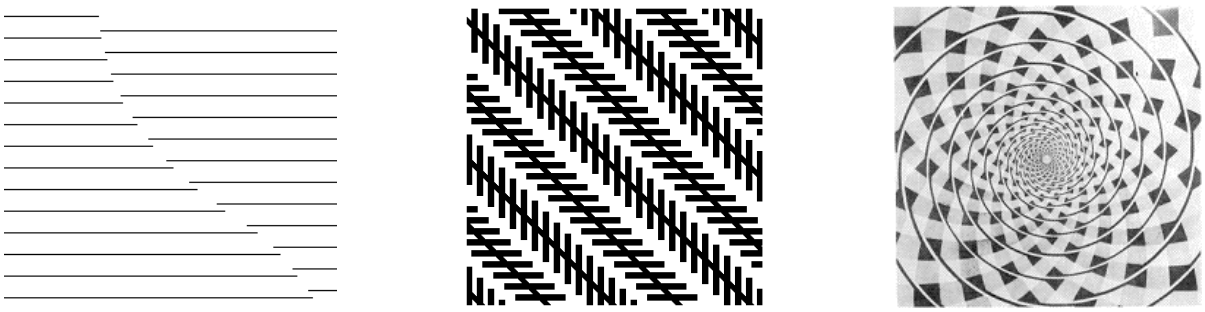


FIG. 1.5 – Phénomènes dus aux interactions spatiales dans l'image. Les contours vrais ou faux (à gauche) jouent par leur orientation un rôle primordial dans certains effets : les diagonales dans l'illusion de Zöllner (au centre) sont parallèles et la spirale de Snearl (à droite) n'est formée... que de cercles concentriques.

résultats assez spectaculaires sur la figure 1.5. Leur explication n'est pas aussi claire que dans le cas de la figure 1.3 à gauche, mais se situerait cependant au niveau de la structure neurophysiologique des voies visuelles où des cellules extrêmement spécialisées dans le traitement des contours ont été découvertes. Mais il existe des illusions notamment sur la dimension des formes (figure 1.4 à droite), pour lesquelles on n'est plus en mesure de donner une explication d'ordre neurophysiologique. Elles correspondraient à des processus mentaux beaucoup plus complexes. Le système visuel ne jouerait déjà plus dans ce cas un rôle aussi passif.

1.2.2 Comportement actif du système visuel

Le rôle actif du système visuel se situe essentiellement au niveau des aires supérieures du cerveau où l'information image arrive après toute une série de transformations et de codages que nous n'avons fait qu'entrevoir. Là, l'information y est rassemblée et synthétisée par des opérations mentales très nombreuses mais encore inconnues. Regardez le trapèze élémentaire de la figure 1.7-A. Suivant le contexte spatial où on le place (B ou C) on en modifie radicalement la perception puisqu'on aura tendance à le voir en position debout ou allongée.

La présence d'éléments de perspective ou de gradients de texture donne en effet de fortes indications sur les distances relatives des objets et permet à l'observateur de mieux structurer l'espace qu'il regarde. Vous constaterez ainsi que les déformations planes dans la figure 1.8 sont finalement interprétées comme un volume caché sous

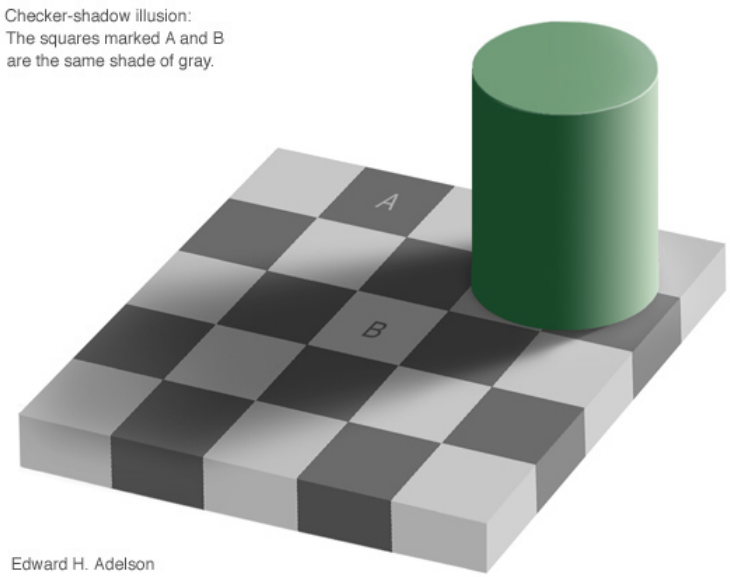


FIG. 1.6 – Dans cet échiquier, les deux carrés marqués A et B ont le même niveau de gris. Plusieurs phénomènes se conjuguent pour renforcer cette illusion. certains relèvent de processus de contraste simultané, d'autres de phénomènes d'interprétation de la scène. © E. H. Adelson.

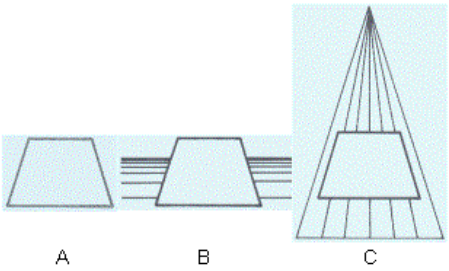


FIG. 1.7 – La perception des formes est modifiée selon le contexte spatial où elle est située.

un damier régulier. Ces indications peuvent aussi agir sur des mécanismes mentaux particuliers comme celui de la constance de la taille et provoquer de fortes illusions géométriques comme la figure 1.9. On y vérifie pourtant que les trois motifs sont absolument identiques et que par conséquent les images rétiniennes qu'ils forment ont la même taille.

Ces exemples soulignent le caractère actif avec lequel le cerveau reconstruit l'information image mais n'indiquent pas la manière avec laquelle il procède pour y parvenir. Comment faisons-nous pour reconnaître les formes, analyser les scènes ? Actuellement ces questions n'ont pas encore trouvé de réponses satisfaisantes.

Quelques approches par le biais de la théorie de l'information ont été tentées, mais semblent être restées jusqu'ici stériles. En fait une constatation toute simple permet de montrer combien il est difficile d'attaquer le problème de cette manière. La photographie 1.10 à gauche n'est rien d'autre que le positif de la photo de l'auteur, dont la figure de droite est le négatif. Or, il est extrêmement difficile de reconnaître une personne sur un négatif sans un entraînement préalable. Et cependant la quantité d'information contenue dans un positif ou un négatif est

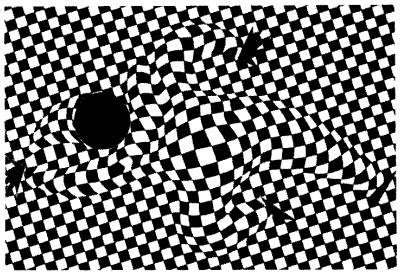


FIG. 1.8 – Rôle des gradients de texture dans l'« Arlequin » de Vasarely (cf. texte).

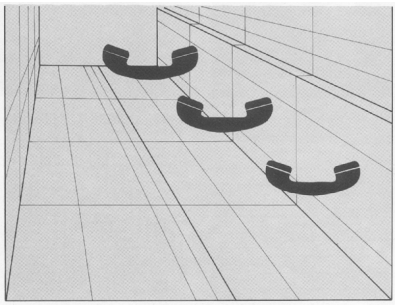


FIG. 1.9 – Illusion géométrique sur la taille provoquée par la perspective.

la même.

L'approche adoptée par les gestaltistes ([Wertheimer, 1944] ou [Gordon, 1998] chapitre 3), très différente, a donné de meilleurs résultats. Pour eux, la perception des signes, des objets ne s'expliquerait pas par un processus de concaténation ou de sommation en série de leurs différentes parties. lis seraient perçus comme un tout. Regardez les 3 segments de la figure 1.11. Lorsqu'ils finissent par se rapprocher suffisamment pour se rejoindre, ils forment une figure triangulaire que l'on perçoit en soi plutôt que comme le résultat de ce rapprochement. Cette construction suivrait un certain nombre de lois indiquées à la figure 1.12. Ces lois poussent à croire qu'il existerait une sorte de principe de simplicité qui dirigerait le cerveau et lui permettrait de limiter au maximum ses efforts pour reconstruire l'information. C'est une idée un peu naïve, mais prenez par exemple le cas des surfaces subjectives de la figure 1.13. Le triangle blanc à droite n'a pas de contour réel pourtant on le perçoit très vivement. Il paraît même être plus lumineux que le fond. Or, qu'y a-t-il dans cette image ? Trois angles et trois disques tronqués. Ces formes paraissent incomplètes au cerveau. Il lui suffit alors de faire l'hypothèse supplémentaire de la présence d'un triangle blanc recouvrant ces formes pour que l'image s'organise plus simplement.

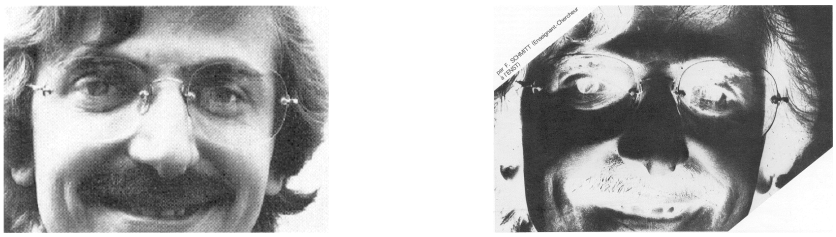


FIG. 1.10 – Positif et négatif de la photo de l'auteur. Vous pouvez vérifier.

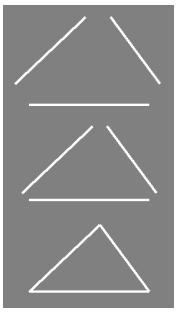


FIG. 1.11 – Le triangle est perçu comme une unité graphique en soi.

Vous constatez que le cerveau ne se contente pas de percevoir que ce qui existe dans le signal. Il n'hésite pas à le compléter si cela l'arrange, c'est-à-dire si cela lui permet de faire son travail de synthèse à moindre frais. Ce que le cerveau rajoute au signal pour effectuer ces opérations constructives dépend bien sûr de l'image mais aussi de la mémoire. Plus la mémoire pourra apporter d'éléments significatifs pour comprendre l'image regardée et moins vous aurez besoin de signal pour percevoir l'information. C'est vrai pour la lecture. Lorsque vous appreniez à lire vous étiez obligé de détailler chaque caractère. Maintenant ceux-ci défilent à toute vitesse devant vos yeux.

Il est des cas où le cerveau peut construire pour une image donnée plusieurs solutions distinctes. Ces solutions s'excluent mutuellement car on ne perçoit que l'une des deux et le passage de l'une à l'autre demande un important effort intellectuel. La figure 1.14 est un exemple célèbre de ces figures ambiguës. Certains reconnaîtront une vieille femme et d'autres une jeune femme fort élégante. Pourquoi ? Il est difficile de répondre. Peut-être les amoureux verront plus facilement la jeune femme alors que les maris tourmentés par leur belle-mère verront plutôt la plus âgée. De nombreux facteurs psychologiques interviennent et la mémoire joue encore ici un rôle prépondérant surtout si on la stimule sélectivement, par exemple en vous racontant au préalable des histoires de vieilles rombières ou au contraire des histoires de belles romantiques.

Inversement avec les images en fausse perspective de la figure 1.15 à gauche, le cerveau ne parvient pas à construire de solution cohérente ce qui provoque même une certaine gêne mentale. À la photo 1.15, à droite, cette gêne diminue et même disparaît. Car il faut réfléchir pour s'apercevoir que les escaliers de cette maison sont d'un genre spécial. C'est le caractère réaliste de la photographie avec ses multitudes de petits détails qui permet au cerveau de s'affranchir de cette construction impossible.

Nous terminerons en signalant que le rôle actif du système visuel se manifeste aussi au niveau de l'analyse physique de l'image, c'est-à-dire dans la manière avec laquelle vous la regardez. Peut-être n'en avez-vous pas vraiment conscience, mais vos yeux bougent beaucoup et pas n'importe comment. Yarbus a enregistré leurs mouvements [Kolers, 1972]. La figure 1.16 en haut montre comment dans une première étape l'axe visuel du regard balaye l'image en suivant très approximativement les contours principaux. Puis dans une deuxième phase, le regard se stabilise sur des points particuliers où l'information est très importante comme les yeux et la bouche d'un visage ou tout autre détail inhabituel. Mais ce qui est particulièrement intéressant c'est de savoir que ce type de « lecture » varie énormément suivant les préoccupations de l'observateur.

Sur la figure 1.16 en bas ont été enregistrées différentes lectures de la photographie figurant en haut à gauche et faites par un même individu. On a simplement posé à ce dernier une question précise correspondant à la scène représentée par cette photographie en lui demandant d'y répondre après quelques secondes d'observation. On peut remarquer alors qu'il va chercher l'information là où elle est la plus significative pour résoudre son problème. Enfin, que tout cela ne vous tourmente pas. Dans deux minutes vous aurez complètement oublié que votre système visuel joue un rôle passif ou actif et vous continuerez à regarder autour de vous comme si de rien n'était.

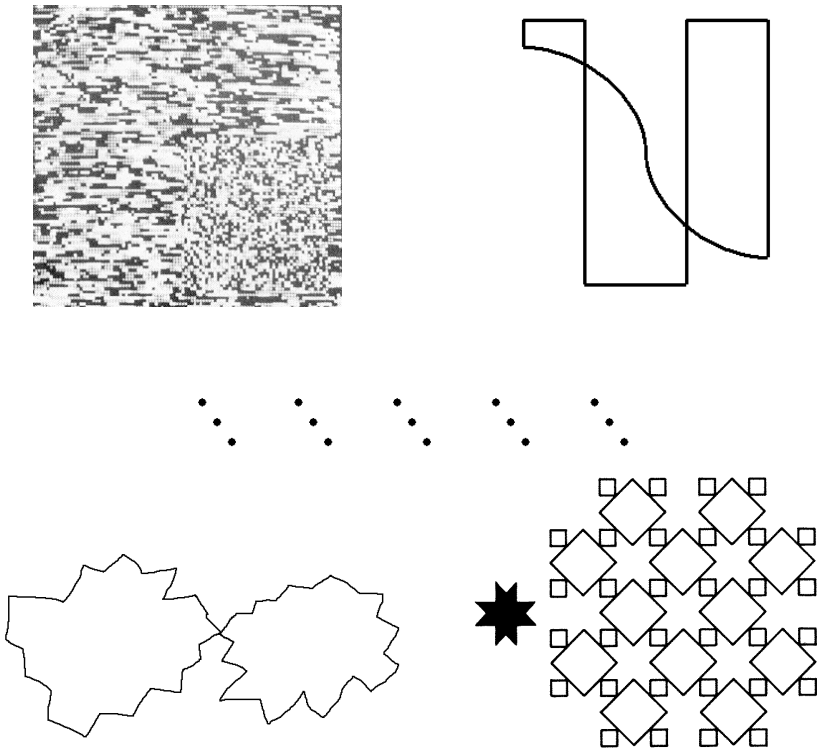


FIG. 1.12 – Lois de la perception (d'après les gestaltistes).
En haut à gauche - lois de similitude : on regroupe les zones ayant des caractéristiques similaires.
En haut à droite - loi de continuité : au lieu de percevoir 3 formes côte à côte, on préfère voir une ligne courbe et une ligne rectangulaire superposée.
Au centre - loi de proximité : on perçoit plutôt 5 triplets de 3 points que 3 lignes de 5 points.
En bas à gauche - loi de fermeture : on perçoit deux formes fermées se touchant en un point plutôt que deux courbes quelconques se croisant.
En bas à droite - le fait qu'on perçoive deux trames de carrés imbriquées, plutôt que la simple juxtaposition de l'élément indiqué en noir est une conséquence de ces lois.

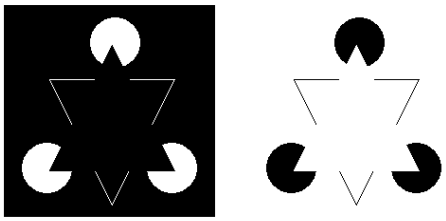
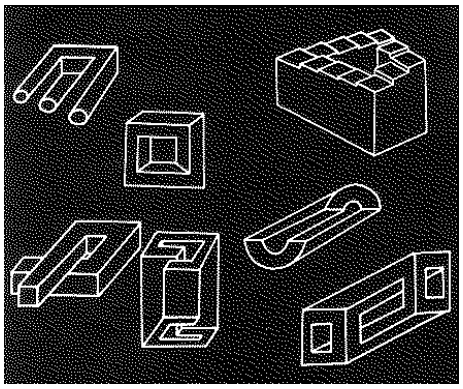


FIG. 1.13 – Deux triangles subjectifs qui n'ont pas de contours (d'après G. Kanizsa, Scientific American, avril 76).

FIG. 1.14 – « Jeune fille ou vieille femme ? » figure ambiguë de R. Leeper.



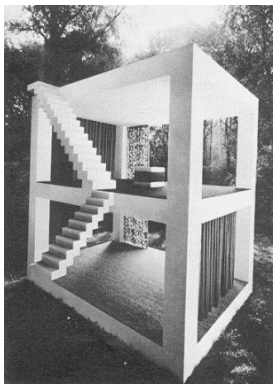
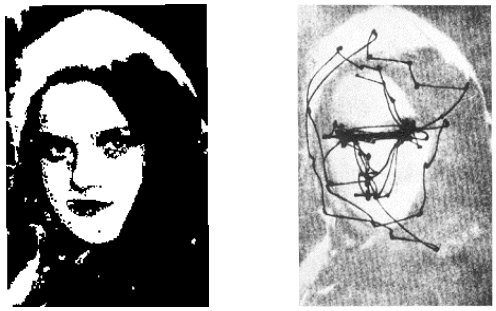


FIG. 1.15 – Quelques constructions impossibles (à gauche), ... qui n'ont pas arrêté cet architecte (photographie A.P. Lamoth), à droite.



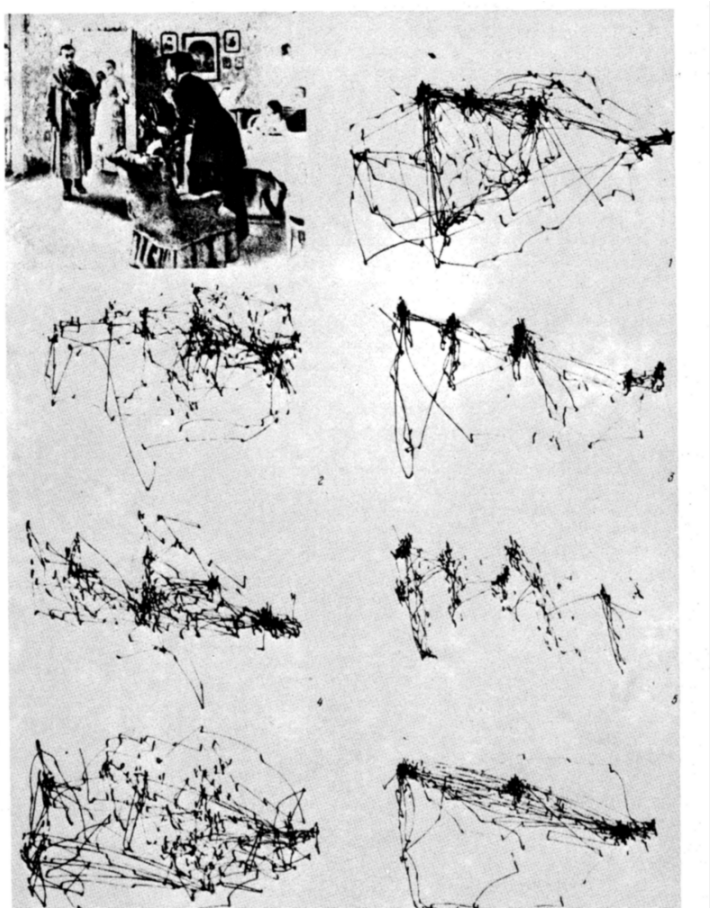


FIG. 1.16 – Chemin que suit l'axe du regard (d'après Yarbus et Kolers [Kolers, 1972]) : en haut, lorsqu'on observe une image ; en bas lorsqu'on demande à l'observateur de se souvenir de diverses choses concernant l'image.