

Agir dans et sur l'espace de travail avec des objets ordinaires

Bernard Conein*

RESUME. Les objets techniques, utilisés tous les jours, lorsqu'ils sont placés dans l'espace, peuvent fournir une aide externe précieuse à l'action sous la forme d'un étayage de la même manière qu'un conseil ou une instruction communiqués à un novice peuvent lui faciliter l'exécution d'une tâche difficile. L'article souligne comment les propriétés spatiales et graphiques inscrites dans les objets et leur placement servent d'aide à l'action. Deux exemples illustrent ce rôle des représentations spatiales : les écrans à touches des billetteries automatiques et les gobelets informationnels des cafés *Starbucks*.

Mots clés : affordance, indices, espace de travail, objets.

ABSTRACT. The everyday use of objects, placed in a workspace, can play the role of a scaffolding or external aid for action control as much as an advice or a verbal instruction can facilitate the accomplishment of a difficult task. The paper underscores how spatial and graphic properties act as action aids. Two examples will illustrate the role of spatial representations for action control: the touch screen in automated ticket delivery and the *Starbuck* drink cup.

Key Words: affordance, cues, workspace, tangible computing, artefacts.

« Nous récupérons nos pensées en retrouvant dans l'environnement les objets qui les représentent. » Norman (1991)

Les objets techniques, utilisés tous les jours, lorsqu'ils sont placés dans l'espace, peuvent fournir une aide externe précieuse à l'action sous la forme d'un étayage (Kirsh, 1999 ; Clark, 1997) de la même manière qu'un conseil communiqué à un novice peut lui faciliter l'exécution d'une tâche difficile. L'étayage (*scaffolding*), selon Clark¹, ne doit pas se réduire à un accompagnement social de l'apprentissage mais inclut les matériaux manipulables et tangibles qui composent l'espace de travail ainsi que l'arrangement de ces matériaux sur une surface et l'ameublement de l'espace par des équipements

* Nice-Sophia Antipolis, courriel : Bernard.Conein@ehess.fr

¹ « Une action est étayée selon l'étendue du rôle joué par les supports externes. Le support peut provenir de l'usage d'outils ou de l'exploitation des connaissances et des aptitudes des autres ; c'est-à-dire, l'étayage pour moi désigne une large classe d'augmentations physiques, cognitives et sociales, et ces augmentations permettent d'accomplir certains buts qui seraient autrement au-delà de nos compétences. » (Clark, 1997 : 194-195).

stables (Lave, 1988). Une des conséquences est que la technologie sous forme d'objets manipulables et d'équipements stables peut devenir un appui pour une connaissance de nature visuelle et spatiale qui ne prend pas la forme de descriptions d'action. Le gouffre entre un plan d'action composé d'instructions et sa réalisation détaillée serait alors comblé par les indications que fournit l'aménagement des objets dans l'espace de travail. L'arrangement de cet espace au moyen d'artefacts utilisés de façon courante suscite des indications pour l'action selon un mode spécifique à partir de représentations spatiales orientées vers l'action (Conein & Jacopin, 1994 ; Conein, 1998). Il est maintenant admis, dans une variété de traditions, que cette connaissance spatiale qui facilite l'accomplissement des tâches quotidiennes n'est représentable ni sous forme d'un programme, d'un plan ou d'une instruction, ni sous forme de représentations conceptuelles propositionnelles.

Nous prendrons deux exemples pour illustrer comment un espace arrangé par des objets techniques contrôle les mouvements d'exécution. Le premier exemple concerne des observations sur la commande de billet de train dans les billetteries automatiques des gares. Le second porte sur la commande de boisson dans les cafés *Starbucks*. Ces deux exemples illustrent le même phénomène : des représentations externes de nature spatiale, inscrites sur les objets, deviennent des appuis pour agir en contexte. Dans les deux cas, les objets servent de rappel (Norman, 1991-1993) ou de point d'entrée (Kirsh, 2001) en contribuant à simplifier les actions, car ils construisent un pont entre nos intentions d'action et les mouvements que nous exécutons.² Cette contribution des objets et de l'espace à l'action en cours résulte d'un couplage dense entre deux partenaires, l'agent et son environnement local familiarisé, chacun contribuant à l'accomplissement de l'action.³ Mais ce couplage par action réciproque n'est pas donné, car il fait constamment intervenir la manière dont l'agent adapte son environnement et utilise les informations, spatiales et visuelles, qui proviennent des objets quotidiens placés dans l'espace. Pour comprendre ce couplage, il faut comprendre non seulement comment des gestes et des mouvements pertinents pour l'exécution se déclenchent à partir d'indices et d'*affordances*,⁴ mais aussi comment les repères visuels et spatiaux sont détectés à partir des diverses façons dont les objets nous apparaissent au moment d'une action.

Cette intervention des objets dans l'action se présente selon plusieurs modalités qui correspondent à des manifestations différentes de leur fonction informationnelle. Un objet technique peut être un support pour des *affordances* (naturelles ou intentionnelles), un réceptacle pour des indices, et une surface pour une instruction écrite. Le même objet physique peut jouer ces diverses fonctions. Il peut aussi être à la fois un soutien informationnel pour une évaluation de la situation et une aide pour des manipulations. Pour cela, il faut

² « Les objets du monde jouent un rôle de signe et il n'est pas nécessaire d'amplifier cette aptitude naturelle pour qu'ils puissent servir de représentations, de symboles. Tout ce qu'il faut c'est un mécanisme associatif pour établir une correspondance entre objets, localisations dans l'environnement et pensées internes. Nous récupérerons nos pensées en retrouvant dans l'environnement les objets qui les représentent. Les objets sont les supports d'une fonction de rappel » (Norman 1988).

³ Comme le soulignent Clark (1997) et Kirsh (1999), les deux parties, agent et environnement, sont interdépendantes et manifestent une causalité réciproque.

⁴ La distinction entre indice et *affordance* intentionnelle est importante pour Kirsh, car elle permet de comprendre que l'espace n'est pas uniquement un espace de contrôle des mouvements mais peut aussi faciliter l'attention, une inférence située et la détection d'une cible. Les indices sont aussi des « invitations à agir », ils ne sont manifestes à l'agent que par inférence.

reconnaître que les fonctions qu'un objet peut porter pour un agent sont multiples : rappel, déclencheur, marqueur et contrainte.

D'où deux questions :

De quelle façon des objets placés dans l'espace permettent-ils d'accorder un programme d'action avec les mouvements corporels qui servent à exécuter ce programme ?

Comment la fonction de contrôle de l'action est-elle réalisée par les propriétés spatiales et graphiques des objets ?

1. LIRE DES INSTRUCTIONS ET APPUYER SUR UN BOUTON

Dans les activités routinières de la vie quotidienne, nous agissons en délibérant peu et sans interpréter chaque symbole sous notre regard. Si dans la nature, l'environnement s'offre comme un ensemble d'*affordances* pour se déplacer et agir avec les objets préhensibles (Gibson, 1979)⁵, dans les environnements équipés comme les bureaux, les ateliers ou les cuisines, les *affordances* naturelles ne suffisent pas car l'espace est constamment altéré par les actions complémentaires d'aménagement des experts pour instancier des indices pour faciliter leur tâche (Kirsh, 1999). Même s'il existe des annotations ou des instructions inscrites sur les ustensiles ou affichées sur des indicateurs, l'attention de l'agent est d'abord engagée dans le mouvement qu'il accomplit avec les objets. Ce sont des indices qu'il détecte plutôt que des symboles qu'il interprète. Les *affordances* et les indices sont des repères pour le contrôle de l'action, ils guident les mouvements et les déplacements dans l'espace de travail. Au cours de l'exécution du geste pertinent, l'intervention du langage semble devenue superflue, un peu à la manière dont la cuisinière chevronnée se passe du livre de recette : même s'il est sous la main, il ne sera utilisé qu'en cas de besoin.

Il semble qu'il y ait un lien entre la manière dont nos actions ordinaires sont accomplies dans l'enveloppement d'un espace familiarisé et le faible degré d'intentionnalité et de réflexivité dans la réalisation de celles-ci. L'interprétation sémantique comme la représentation conceptuelle des actions joueraient un rôle mineur dans l'exécution des gestes routiniers au détriment d'autres aptitudes de nature visuelle et spatiale.

En conséquence, les études sur les espaces de travail (Suchman, 1987 ; Hutchins, 1995 ; Kirsh, 1995) s'éloignent d'un modèle de l'action accordant un rôle prédominant à la fois aux représentations sémantiques (ou aux croyances) et à la délibération sur les choix. Ils cherchent à montrer de quelle manière l'information spatiale, inscrite dans l'environnement et les objets, parvient à l'agent en temps réel, au moment même de l'accomplissement de l'action. Mais ces travaux récusent aussi l'idée que seul l'environnement physique est susceptible par lui-même de soutenir nos actions, car l'espace de travail se présente comme constamment modifié par la présence des artefacts, la localisation des objets, la cohérence des équipements et la familiarité des habitats. L'environnement ne se réduit donc pas à une structure d'*affordances* naturel-

⁵ Les *affordances* visuelles d'un objet (Gibson, 1979) contrôlent les mouvements d'exécution de base sous forme de couplage quasi-automatique entre perception et action à travers la prégnance d'une forme ou d'un aspect de l'objet : la forme d'un verre engage à le remplir, celle d'un bâton à le saisir, celle d'une pierre à la lancer. Les indices sont plus inférentielles, leur fonction d'indication se réalise à cause des certaines propriétés spatiales qui ont construites par l'aménagement de l'environnement.

les, s'il est aménagé par le dépôt d'*affordances intentionnelles*⁶ et par l'instanciation des indices au moyen des actions complémentaires de stabilisation. Ces indices agissent alors comme des points d'entrée qui sont des invitations à agir : « un point d'entrée est une structure ou un indice qui représentent une invitation à agir – à s'engager dans un nouveau chemin ou dans un espace informationnel » (Kirsh, 2001).

Cette variation dans le partenariat agent-environnement conduit à s'interroger sur la contribution des objets à la stabilisation de l'environnement, et sur les frontières qu'ils tracent selon leur placement. Ainsi les indices que nous déposons au fur et à mesure que nous agissons en son sein font partie de l'environnement bien qu'ils soient le résultat de sa stabilisation.

La couleur verte du bouton de la photocopieuse *Rank Xerox* est un indice pour enclencher l'action de copier. L'*affordance* du bouton (appuyer) ne peut être détectée si on ne sait pas où il se trouve. Comme le montre l'exemple de la conversation entre deux novices autour d'un photocopieur intelligent, enregistrée par Suchman (1987), l'environnement peut être un obstacle à l'action s'il se présente selon deux modalités au départ non coordonnées : une instruction sur un écran et un système de commande.

A :	« Appuyez sur le bouton de démarrage » Où est le bouton de démarrage ? ((tourne autour de la machine et ensuite regarde le tableau d'affichage))
B :	((Pointe sur le bouton de démarrage)) Démarrage ? Regarde juste là
A :	Ici ((rire))
B :	D'accord

Le programme d'instructions affichées sur l'écran du système d'aide est d'abord le centre de focalisation des deux novices qui voient la photocopieuse pour la première fois. Sa lecture reste un passage obligé pour agir. Quant à l'indice qu'est la couleur du bouton de démarrage, il ne devient pertinent qu'après la coordination entre A et B. Dans cet exemple, la relation avec les objets et l'espace de travail prend au départ la forme de ce que Norman (1991)⁷ appelle une double coordination : A est coordonné avec le menu et B avec l'environnement de la tâche. A lit l'instruction (« appuyez sur le bouton de démarrage ») mais, face à un environnement technologique non familier, il tombe sur un gouffre informationnel car l'instruction ne représente pas l'information spatiale pertinente : la localisation du bouton de démarrage. B coopère avec A en se focalisant sur l'environnement de la tâche, l'endroit où se

⁶ Tomasello (1999) parle d'*affordances* intentionnelles pour les ressources dans le contrôle du mouvement qui viennent d'un artefact. Dans ce cas, les *affordances* sont conçues intentionnellement par un ingénieur concepteur ou par l'agent quand il aménage son environnement (Hammond *et al.*, 1995).

⁷ Norman prend l'exemple de la *checklist*, dans la conduite d'avion. Celle-ci lorsqu'elle est consultée implique une dualité des interactions : une interaction avec l'artefact et avec le *cockpit*.

trouve le bouton, et quand il trouve son emplacement, il pointe son doigt. Du point de vue de l'interaction, la scène ne se réduit pas à une simple interaction conversationnelle, car elle se présente comme une scène d'attention conjointe tripartite entre trois composants : A, B qui s'orientent progressivement vers C (le bouton vert), qui devient ensuite un objet commun de focalisation.

Par contre, lorsque l'agent est couplé à son environnement par l'intermédiaire d'objets, l'interaction n'a plus la forme ni d'une double coordination (vers le menu et vers l'environnement de la tâche), ni d'une co-orientation triadique (2 agents + un environnement) mais d'un couplage direct agent/environnement. Pour un utilisateur routinier de la photocopieuse, l'espace du problème a changé car l'interaction avec l'environnement comme la scène interactionnelle ont changé. D'abord l'utilisateur n'est plus en situation de recherche d'information et il n'a besoin ni de l'assistance d'autrui, ni de lire les instructions car l'environnement spatial est devenu pour lui son meilleur système d'aide. En effet l'utilisateur ne lit plus les instructions même s'il peut jeter un coup d'œil sur l'écran avant d'appuyer sur le bouton vert ; il traite alors l'aspect graphique de l'inscription comme un pointeur pour appuyer sur le bouton. Il peut réduire l'inscription à un indice, la couleur verte du bouton lui servant de moyen pour savoir à quel endroit appuyer. L'information visuelle et spatiale dans un environnement familier sert de guidage lorsque les informations linguistiques présentes sur l'affichage sont superflues. Le regard porté alors sur l'écran n'est plus un regard de lecture littérale mais de détection de repères.

Un autre cas illustre des phénomènes similaires : des informations sémantiques peuvent être traitées comme des indices si les agencements spatiaux et visuels sont suffisamment saillants pour l'utilisateur.⁸ J'ai observé à la vidéo les utilisateurs des billetteries automatiques dans les gares. La SNCF a développé des interfaces composées des touches étiquetées sur écran. Ce système tente de mieux coordonner les informations sur l'écran avec les boutons d'enclenchement des automates en utilisant le même objet comme indicateur et déclencheur d'une commande. Cette interface utilise la surface de l'écran pour enclencher le mouvement de manipulation des touches, la touche étiquetée servant à la fois d'indicateur et de commande à manipuler. Ainsi, ce dispositif tente de résoudre le problème de la double coordination en utilisant le même support spatial pour deux actions : lire l'information sur le trajet et presser une touche.

Dans cet exemple, l'écran à touches augmente les propriétés de contrôle sur l'action.⁹ D'abord, la distance spatiale entre menu et commande est supprimée, l'écran à touches conjugue deux fonctions : indicateur et contrôle. Ensuite, le contenu sémantique des informations dans le menu est schématisé par les noms de destination disposés sur la surface de l'écran. Enfin, l'arrangement spatial de l'écran (disposition, groupement, figure/fond) focalise sur l'information sémantique la plus schématisée (le nom de la destination sur la touche).

Si l'on examine maintenant la relation du menu aux boutons du point de vue du contrôle de l'action, l'agencement des touches sur l'écran permet de considérer la surface comme un tableau de commande. La disposition spatiale

⁸Marc Breviglieri et moi avons filmé dans plusieurs gares des utilisateurs prenant leur billet aux billetteries automatisées.

⁹ Bien qu'il s'agisse d'une interface graphique, la possibilité de manipuler « directement » l'écran redonne à la manipulation une fonction plus effective.

des touches étiquetées sur l'écran présente en effet un taux élevé de contrôle qui exploite au maximum la dimension spatiale et au minimum la dimension conceptuelle. Comme dans l'exemple de la photocopieuse, ce n'est pas le contenu descriptif de l'instruction qui importe, mais la disposition spatiale et en particulier la proximité entre l'information sur l'action et le contrôle de la manipulation.

Examinons de plus près comment on obtient cet effet de contrôle par l'environnement. La touche étiquetée présente sur l'écran le nom d'une destination. L'écran présente les contenus par degré d'abstraction : d'abord le but général de l'action, « Billet Banlieue » ; puis un verbe d'action de forme directive « Choisissez la destination » ; enfin, le sous-but « Aéro.Orly » qui spécifie un type de destination et constitue en même temps la cible du geste de pression puisque l'utilisateur doit appuyer sur l'espace encadrant le nom.

La représentation conceptuelle du choix de destination est simplifiée pour que la disposition spatiale sollicite la connaissance visuelle et spatiale en agencant les représentations selon trois régions. Une région supérieure identifie l'objet et le but général (Billet Banlieue), une région intermédiaire indique une action de sélection (Choisissez) et une région centrée sur l'écran affiche trois sous-buts (destination, classe + tarif et nombre de titres). Les touches étiquetées se présentent comme des rectangles encadrant les spécifications du contenu des buts (une destination, une classe et un tarif, un nombre), aucune instruction n'évoque le geste de pression à exécuter pour enclencher la touche. Si les deux premières régions de l'écran sont pertinentes pour une lecture séquentielle des instructions, elles deviennent facultatives pour un utilisateur régulier qui utilise les propriétés spatiales (localisation des touches, leur distance relative) pour enclencher la commande de billet¹⁰.

Billet Banlieue

CHOISISSEZ

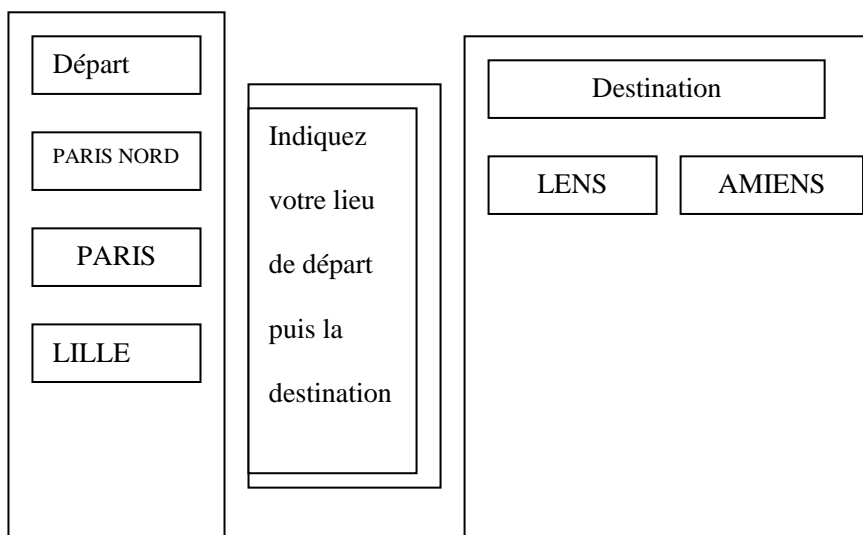
1. La destination

/Aéro.Orly RER C/	/Epinay-Villet./	/Pierrefitte-St./
/Aréo.Orly Val/	/Garges-Sarcelles/	/Plaine-Stade-Fr/
.....		

Dans cette interface, la surface de l'écran a donc un double rôle, à la différence du menu de l'écran de la photocopieuse. Elle sert à la fois de zone manipulatoire pour presser les touches et d'affichage de l'information sur les buts. La zone manipulatoire est réduite aux rectangles qui encadrent chaque nom de destination.

Pour d'autres affichages (Grandes Lignes) les touches sont réparties en deux tableaux : un tableau à gauche exposant sous l'en-tête « départ » les villes principales de départ et un à droite exposant sous l'en-tête « destination » les villes principales d'arrivée. Entre les deux lignes du tableau est placée une instruction « Indiquez votre lieu de départ puis la destination ».

¹⁰ C'est moins la régularité de l'action qui prédomine que les ressources fournies par l'agencement des représentations spatiales et leur combinaison avec la manipulation des touches (cf Pacherie, 2000).



Si l'utilisateur prend un trajet typique avec deux gares principales, il peut utiliser une routine d'usage des touches par une pression à gauche et une pression à droite sans arrêt entre les deux mouvements. Ces routines d'usage s'appuient sur une interprétation des relations spatiales. Le contenu de l'instruction n'a pas besoin d'être traité car les buts peuvent être traités comme des cibles d'un mouvement.

L'interprétation spatiale est mise en œuvre par compréhension indexicale :

i. les énoncés centraux (destination, classe, type de billet) sont indexés à des objets manipulables dans l'environnement (par exemple : la destination sur la touche),

ii. les structures prégnantes sur l'écran (encadrement, liste, ligne) établissent des correspondances de façon à associer étroitement cible et mouvement. La disposition spatiale des inscriptions et des touches contrôle les mouvements de pression des touches. Dans le cas de touches étiquetées, la localisation permet d'associer le symbole « Paris » à une touche sur écran.

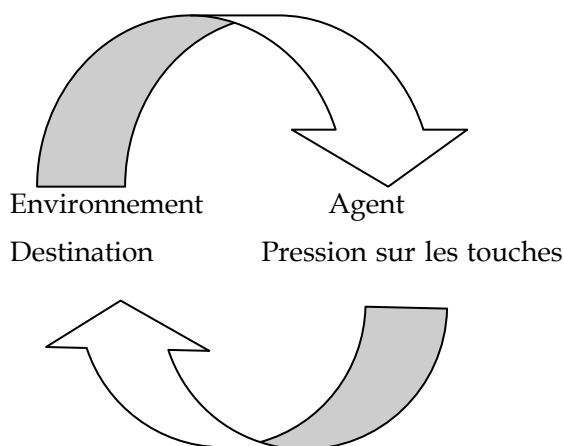
La touche étiquetée implique que la même représentation serve de support à l'évaluation de l'information et au contrôle de l'exécution. L'écran constitue une zone manipulatoire.¹¹

Certaines inscriptions sont susceptibles de constituer un point d'entrée pour un geste au moyen soit d'une inférence, soit d'une perception directe. Il y a en effet une nuance entre voir la destination et ensuite appuyer sur la touche et voir et appuyer immédiatement, même si dans les deux cas, l'action ne prend pas la forme d'une séquence composée d'une phase délibérative de lecture de l'inscription avec interprétation du contenu, suivie d'une phase de décision et d'exécution du geste.

¹¹ J'emprunte la notion de zone manipulatoire à G.H.Mead (*Philosophy of the Present*) qui constitue l'espace de la tangibilité des objets sous la main et à portée de main (cf. Conein, 1997 ; Conein & Jacopin, 1999).

Lorsque l'interprétation est inférentielle, elle peut impliquer que le regard traite deux régions de l'écran sans pour autant passer par un temps délibératif. L'énoncé « Choisissez la destination » est plutôt conçu comme un pointeur sur la touche étiquetée encadrant « Aéro. Orly Val ».

Lorsque l'interprétation est spatiale, les appuis à l'interprétation sont des relations spatiales. La modalité de l'interprétation se fait alors sous forme d'un couplage perception/action. La pression directe sur la touche étiquetée « Aéro. Orly Val » n'implique plus que l'énoncé sur le choix de la destination soit traité car une seule région est focalisée. La focalisation sur l'emplacement de la touche étiquetée suffit dans ce cas à déclencher la pression sur le rectangle qui encadre le lieu de destination :



COUPLAGE

C'est seulement dans ce second cas que la connaissance visuelle et spatiale devient prédominante pour la manipulation de la touche. La touche devient la surface de visualisation de l'information, et en même temps du contrôle du mouvement, comme dans l'exemple de l'objet-symbole mentionné par Norman.¹²

Le caractère indexical du contenu de l'instruction pour le passage à l'exécution se marque de la manière suivante. Le contenu de l'inscription (choisissez la destination) ne donne aucune indication littérale sur le comment-faire, sur comment choisir la destination, ni où faire (à quel endroit), la seule instanciation littérale du contrôle étant le 1. qui indique l'ordre séquentiel des mouvements de pression.

Par contre, la modalité graphique sous forme d'un tableau comprenant une liste de touches avec des destinations donne des informations spatiales sur la coordination de la vision et du toucher.

L'interprétation spatiale n'est pas métaphorique, elle repose sur l'ancrage physique dans l'environnement et s'estime en termes de disposition, de dis-

¹² Un objet symbole est un objet tangible qui suppose que l'information pour évaluer l'action et l'information pour le contrôle de l'exécution se trouvent au même endroit (Norman, 1991-1993).

tance relative, de proximité et d'orientation. C'est le contenu spatial de la disposition des inscriptions qui permet de construire une routine d'exécution qui focalise immédiatement l'attention sur la touche de destination en associant un sous-but (le nom de la destination souhaitée) et une touche. L'exemple des touches étiquetées sur un écran montre qu'un format graphique sur un support spatial (rectangle placé sur une ligne insérée dans un tableau), en encadrant la cible (la touche), peut servir d'appui au mouvement.¹³

2. LIRE LE CODE ET REMPLIR LE GOBELET

Notre second exemple porte sur l'usage des gobelets codés dans les cafés *Starbucks*. Il se rapproche du premier cas par plusieurs aspects. Comme le premier cas, une tâche routinière, accomplie tous les jours, peut être modifiée par l'usage d'un objet unique et d'un arrangement de l'espace pour placer cet objet. C'est aussi le même objet qui sert de surface informationnelle et de surface manipulatoire. Par ailleurs, la commande verbale et la transaction sociale (un billet au guichet, une boisson au comptoir) sont transformées en une tâche d'évaluation et de manipulation motrice réalisée par un agent unique. Dans les deux cas, l'action de commande d'un produit, au lieu d'être implémentée sous forme d'une commande verbale adressée à un agent, est réalisée sous forme d'une succession de gestes contrôlés par un codage sémantique minimal.

L'usage du gobelet codé est dépendant de certains contextes de commande. Dans les cafés *Starbucks*¹⁴, ce n'est que lorsque la préparation de la boisson est complexe ou qu'il y a de l'attente que les serveurs utilisent un gobelet en carton ou en plastique transparent pour inscrire leur commande sous forme d'un code convenu.

En effet, à chacune des étapes, différents agents (le donneur d'ordre, le préparateur, le client) utilisent le même ustensile pour faire des choses différentes sur trois espaces différents : le comptoir de commande, l'espace de travail de la machine à café et la table en salle. L'objet gobelet se présente dans cette situation de trois manières : comme *surface* d'inscription d'un code, comme *réceptacle* pour recevoir la préparation, et comme *tasse* pour boire. Ce ne sont pas les mêmes aspects de l'objet qui sont concernés dans chacune des activités, mais c'est le même objet saisissable et portable qui sert de support (Marr 1982).

Le codage de la commande se manifeste par des lettres ou des signes inscrits sur six carrés blancs. Lorsque l'ensemble des inscriptions est porté sur le gobelet, le script est traduisible pour le préparateur qui compose et sert la boisson. Le client règle la boisson pendant que le préparateur la compose.

Les six carrés blancs imprimés sur la surface du gobelet en carton sont les options dans le choix des ingrédients pour composer la boisson en fonction de la commande.

¹³ La disposition spatiale entre le lieu de départ et le lieu de destination sur le même support permet aussi que la première action (choix d'une ville de départ) soit immédiatement suivie de l'action suivante (choix d'une destination).

¹⁴ Nous avons réalisé, avec David Kirsh, plusieurs observations sur la commande de boisson dans un café *Starbucks* à Cardiff dans le canton de San Diego en octobre 2004.

Decaf = décaféiné

Shots = 1 ou 2 doses

Syrup = type de sirop

Milk = lait (+ ou -)

Custom = spécial

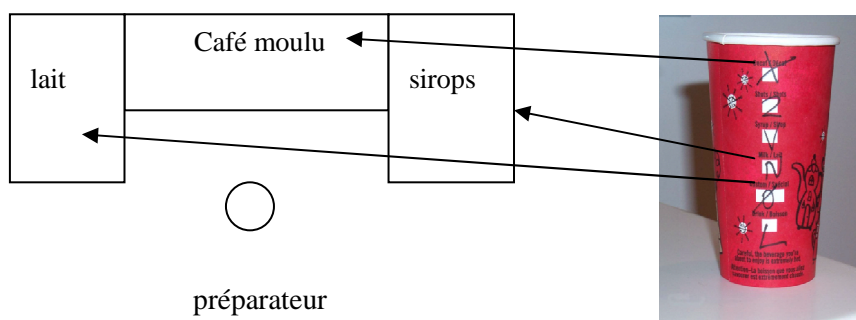
Drink = genre



Cette schématisation de la composition de la boisson permet une réduction importante du temps de réalisation de l'action de commande et de sa préparation, mais aussi, comme dans l'exemple de la photocopieuse, on assiste à une transformation des actions, des mouvements corporels et de la scène de la transaction.

La schématisation de la commande, sa réduction à un code, guide la préparation de la boisson sur l'espace de travail, car elle réduit au minimum l'information sémantique et amplifie au maximum la fonction de contrôle pour la préparation.

De quelles ressources, le préparateur a-t-il en effet besoin pour préparer la boisson ? Il a besoin d'une mémoire externe, pour se rappeler des composants de la boisson commandée, des objets et des ingrédients à portée de main, et d'un espace de travail qui met ces ressources à portée de main. L'économie de temps dans l'exécution de la commande de boisson par la réduction des transactions sociales n'est pas séparable d'une modification de la réalisation motrice de la composition de la boisson. Cette modification est obtenue par la mise en correspondance entre l'espace d'affichage des inscriptions et l'espace de composition de la boisson avec les ingrédients, afin que chaque item mentionné sur le gobelet corresponde à son placement dans l'espace manipulateur.



Pour comprendre la façon dont s'établit la coordination temporelle entre ces différentes phases et les actions, il faut repartir du système fonctionnel de la tâche, et évaluer comment sa forme d'implémentation dans l'environnement modifie les actions au sens étroit. A une certaine échelle, il semble que ce soient les mêmes actions qui sont réalisées au moyen d'un gobelet codé ou d'une commande verbale, alors qu'à une autre échelle, l'implémentation motrice étend l'extension des actions en intégrant les mouvements corporels.¹⁵

Si l'on simplifie le système fonctionnel de la tâche sous forme d'un programme d'actions indépendant de son implémentation, de sa réalisation motrice et du canal de transmission de l'information sur la commande (modalité verbale ou modalité spatiale), quatre grandes fonctions à réaliser se présentent : énoncer la commande, noter la commande, payer et encaisser, préparer la boisson.

Dans sa forme standard d'implémentation de ce système, chaque transaction est allouée à un espace distinct et se réalise de façon séquentielle :

- implémentation de la commande par *une communication verbale* entre le client et le serveur au comptoir, sous forme d'une séquence salutation/requête du client/inscription de la commande/paiement ;
- implémentation de la préparation de la boisson *dans les gestes de manipulation des boissons*, derrière le comptoir, sur la machine à café par le serveur;
- implémentation de la prise de boisson *dans les gestes de saisie* du gobelet par le client soit sur un comptoir, soit sur une table.

Il existe donc deux phases principales avant la prise de boisson : *la prise de commande* sous forme d'une transaction avec le client et *la préparation de la boisson* sous forme d'une transaction entre les serveurs.

Comment le système d'encodage des informations sur la boisson dans le gobelet modifie-t-il la séquentialisation des actions et les mouvements corporels du préparateur de la boisson ? Le gobelet codé intervient en effet dans la mise en phase de la prise de commande avec la préparation de la boisson. Le gain de temps vise d'abord ces deux actions. La commande est formatée par le système de codage, qui est lui-même une schématisation du menu. Ce système modifie la structure de coût de l'activité, mais aussi la transmission de

¹⁵ Je reprends à Corazza & Dokic (1993), la distinction entre action au sens étroit et action au sens large. Les actions au sens large incluent une référence aux circonstances de l'action.

l'information sur la boisson et la localisation spatiale, la scène attentionnelle, et donc la transaction entre le preneur d'ordre et le préparateur. L'introduction du gobelet codé change le support de l'information sur la commande et sa représentation pour l'utilisateur, en substituant à une production verbale une manipulation guidée par des indices et un schéma graphique pour coder les actions.

Les effets du codage graphique de la commande sur le gobelet se présentent à deux niveaux. Le gobelet modifie les séquences motrices et visuelles en modifiant à la fois la scène attentionnelle, donc la transaction, et son implémentation motrice.

Concernant la scène attentionnelle, le preneur d'ordre n'est plus en interaction conversationnelle dyadique avec le préparateur, mais il construit une scène d'attention où le gobelet devient l'objet d'attention unique du préparateur. Le preneur d'ordre pose en effet le gobelet codé sur le haut de la machine à café à la suite des autres gobelets en attente d'être remplis. Ensuite, le préparateur prend le gobelet selon l'ordre de placement. Enfin lorsque le préparateur interprète le codage des actions, il le fait à l'intérieur de l'espace de travail de préparation de la boisson dans le cours d'action de la préparation de la boisson. Il peut ainsi mettre en correspondance les composants d'un programme d'action schématisé avec des mouvements de manipulation des objets (prendre un gobelet, le remplir avec du café, du sirop, du lait, etc...)

Concernant l'implémentation motrice, il est intéressant d'évoquer une idée de Goody (1993-1994) sur la modification de l'implémentation du comportement lorsque le canal de communication change. Lorsqu'on lit des inscriptions graphiques, l'agent est en contact avec un support visuel et spatial. Par contre lorsqu'il entend des énoncés prononcés, l'oreille entend ce que dit une bouche. Pour Goody, la scène modifie la manière dont l'information se présente dans l'espace, car le codage graphique mobilise l'œil, un support spatial et la main, alors que la parole mobilise l'oreille et la bouche. Lorsque le canal est verbal, le corps est momentanément détaché de la manipulation, alors que lorsque le canal est visuel et que le support spatial sert à la fois à voir et à manipuler, la main est en prise sur le même objet.¹⁶

La conception de l'action qui met l'accent sur la description de l'action ne prend donc en compte ni l'implémentation des actions dans des gestes et des mouvements corporels, ni le rôle de l'espace comme surface pour déposer des indications sur l'action en cours¹⁷. Elle tend ainsi à considérer que la structure fondamentale de l'action n'est pas modifiée par la présence des objets.

3. LA FONCTION DES OBJETS DANS L'ACTION

De ces deux exemples, on peut tirer quelques leçons sur le rôle des objets dans la stabilisation de l'espace de travail. Ces leçons prolongent en partie l'idée de départ de Suchman : les programmes d'action ne sont pas orientés vers la réalisation de l'action, ni vers le mouvement pertinent, ni vers l'orientation du corps et des objets dans l'espace. Ils sont une tentative de dé-

¹⁶ « La fonction de l'écriture est de donner une information auditive sur un support visuel et donc spatial. Le canal de communication passe ainsi de l'état auditif à l'état visuel. On entend la parole et on voit l'écriture ; on parle avec la bouche et on écoute avec les oreilles ; on écrit avec la main et on lit avec les yeux » (Goody, 1994 : 195).

¹⁷ Ces remarques ne sont pas propres à la cognition distribuée, Donald Schön a montré qu'un studio d'architecte constitue une ressource pour apprendre la réalisation d'un plan car les aptitudes clefs sont fortement reliées aux outils et aux matériaux disponibles dans le studio.

crire l'action comme un plan à suivre. Mais leur contenu descriptif n'est pas « ancré » dans l'environnement même lorsqu'il est inscrit sur un support spatial comme un écran. L'information indiquée est à la fois décontextualisée et difficilement contextualisable car la tentative de la contextualiser pour tous est vaine dans la mesure où sa pertinence est intrinsèquement locale et orientée vers l'agent. Suchman oppose l'échec du contrôle de l'action par l'instruction au contrôle de l'action par la dynamique des interactions.¹⁸ Si le plan ne contrôle pas l'action, c'est qu'une description linguistique d'action, comme une instruction affichée ou une commande verbale, ne contrôle ni la motricité, ni la vision, c'est-à-dire les moyens d'exécuter une action de base comme « appuyez sur le bouton de démarrage ». Le travail d'ancrer une représentation linguistique dans l'environnement passe par sa mise en phase avec des mouvements, des placements et des trajectoires (Livet, 2000). C'est la stabilisation de l'environnement qui permet cette mise en phase en utilisant les représentations spatiales comme moyen de contrôle.

Se posent donc trois problèmes distincts qui entretiennent un rapport étroit : celui de la description qui intervient effectivement dans l'action, celui du contrôle de l'action et enfin de la tangibilité de l'objet lorsque le même objet sert à lire et à manipuler.

La description des aspects de l'objet

Le premier problème concerne les représentations des aspects de l'objet qui comptent pour la réalisation de l'action. Il est intéressant ici de convoquer un commentaire d'un analyste de la vision comme Marr (1982) sur la description des objets. Selon lui, chaque objet peut recevoir plusieurs descriptions (descriptions multiples) selon les scénarios d'action qui sont pertinents dans la situation. La représentation multiple de l'objet tient à un rapport particulier qui s'instaure dans l'action, entre la description visuelle des aspects de l'objet et la sélection de ses propriétés fonctionnelles. Lorsque nous avons besoin d'un objet pour réaliser une action, ce n'est pas sa forme, ni sa désignation qui importent mais les aspects de l'objet. Le même objet ne nous apparaît pas de la même manière si on se focalise sur un de ses aspects car, dans ce cas là, il nous apparaît selon un mode particulier. Marr prend comme un exemple un journal. Un journal peut nous apparaître selon trois modes : comme papier si on s'en sert comme combustible, comme support de lecture si on le lit, comme arme si on veut écraser une mouche qui bourdonne. Loin que l'interface entre les propriétés sémantiques de nature fonctionnelle de l'objet se fasse au niveau de sa forme, c'est-à-dire au niveau terminal du processus visuel (représentation volumétrique 3D), Marr envisage qu'en contexte d'usage de l'objet, l'interface entre ces propriétés et les représentations perceptuelles intervienne à un niveau intermédiaire, celui de l'aspect (représentation bi-dimensionnelle 2 1/2 D)¹⁹. Il introduit entre la désignation de l'objet et ses propriétés sémantiques un niveau intermédiaire de représentation qu'il appelle « descriptions brutes », qui a des implications sur le rôle des représentations spatiales dans l'action. Ce niveau intermédiaire constitué par l'aspect de l'objet jouerait un rôle central dans la sélection du contexte lié à une action.

¹⁸ Il faut rappeler que, pour Suchman, l'interaction générique est l'interaction dyadique conversationnelle (Conein & Jacopin 1994).

¹⁹ Pour Marr (1982), le processus visuel passe par trois étapes : esquisse primaire, esquisse 2 1/2 D et modèle 3 D. L'analyse bi-dimensionnelle joue un rôle important dans la détection de la forme.

Tout objet physique est susceptible de générer simultanément plusieurs descriptions, qui chacune spécifie un des aspects de l'apparence de l'objet. Le traitement visuel fournit alors une première description « brute », qui a pour fonction de désactiver les descriptions virtuelles concurrentes. Les informations contextuelles joueraient un rôle central dans la sélection de cette première description. Comme le souligne Marr « nous ne savons généralement pas d'avance quel aspect d'un objet ou de l'action est important ; il s'ensuit que la plupart du temps un objet est susceptible de plusieurs descriptions brutes ayant chacune une forme différente ».

Starbucks est un adepte de la théorie de Marr dans son modèle du gobelet car il utilise la possibilité de susciter à partir du même objet plusieurs scénarios différents d'action selon la tâche en cours (prendre une commande, préparer la boisson ou boire). Le gobelet est disponible pour recevoir du liquide, boire mais aussi pour préparer la boisson. Lorsque je bois, je ne m'intéresse plus aux inscriptions mais à la saveur du liquide. Le caractère brut de la description tient à ce qu'elle évoque des rôles et des buts sans passer par des représentations fines de l'objet.

Le problème du contrôle

Le contrôle des mouvements d'exécution implique que la localisation des objets dans l'espace comme les moyens d'exécuter l'action soient identifiables par l'agent sans coût cognitif excessif. Or, il existe bien un langage spatial qui mentionne les localisations, les groupements et les partitions²⁰. La solution ne serait-elle pas de compléter le contenu des descriptifs des instructions dans les menus en les contextualisant au moyen d'une description spatiale ? Une instruction serait suffisamment détaillée si elle était accompagnée d'une description spatiale du genre : « appuyez sur le bouton vert de démarrage placé au milieu du photocopieur ». Or comme le souligne Dokic (1999), on se trouve ici devant un paradoxe : plus on fixe le détail de l'exécution de l'action avant sa réalisation (mouvement, placement des ustensiles, orientation du corps etc...), plus l'agent est surchargé de représentations trop encombrantes pour agir sur le champ. Par contre plus, les représentations antérieures sont schématiques et simplifiées, plus l'environnement extérieur est susceptible de jouer son rôle de complémenteur au moment de la réalisation des mouvements nécessaires à l'action, comme dans l'exemple de l'écran des billetteries.

Le lien entre vision, intention et mouvement se réalise en fait dans le processus en ligne d'exécution de l'action (Livet, 2000) car la séquentialité entre intention dans l'action et exécution du mouvement se situe à un niveau différent de l'abstraction fonctionnelle du programme d'action. Livet conçoit le cours de l'action située dans la rectification compensatrice et correctrice des mouvements corporels en fonction de ce qui est perçu.²¹ Quant au plan programme, il reste un agencement de buts à une échelle supérieure : « un plan ne détermine nullement les actions de base, il implique simplement les conditions qu'elles doivent satisfaire. Ce n'est donc pas du tout un modèle de l'action mais un modèle de la planification des conduites complexes »

²⁰ Certains mécanismes visuels concernent plus directement les informations spatiales, comme les groupements, les partitions ou la relation figure/fond. Ces mécanismes sont aussi liés à la manipulation des objets ou à l'exécution d'un mouvement (Ullman, 1996).

²¹ Selon Livet, le programme d'action reste un agencement de buts à une échelle supérieure qui ne détermine pas les actions de base (Livet, 2000 : 234).

(Livet : 234). L'instruction ne contrôle pas l'action car elle ne spécifie ni les mouvements, ni les orientations, ni les localisations. Les représentations qui jouent un rôle effectif dans l'action sont orientées vers des objets réels (Ballard et al., 1997) car la seule manière d'enclencher les mouvements pertinents est de configurer l'environnement physique pour qu'ils suscitent des contenus non conceptuels (spatiaux et perceptuels) qui contrôlent l'exécution des mouvements.²²

La tangibilité de l'objet

L'impact de l'objet dans l'espace de travail vient de ce qu'il accorde la main et l'œil d'une manière telle que la détection de son aspect ne s'accompagne pas d'une lecture des symboles et que donc la main n'est pas en attente de trouver sa cible. Le couplage avec l'objet au moyen des représentations spatiales suppose ce lien entre la perception des aspects et le mouvement. Les objets techniques suscitent des pensées mais sous une forme directement liée à l'action. C'est pour cette raison que ces pensées sont indexicales et qu'elles ne sont pas représentées par des formats publics objectifs, graphiques ou verbaux. Leur dimension faiblement publique et commune provient de leur nature indexicale car la manipulation des objets tangibles reste le point de départ naturel de nos interactions avec un espace de travail. Comme l'ont souligné à la fois Agre (1997) et Clark (1995), si nos pensées sur l'action et sur les objets sont indexicales, et donc implicites, égocentrées et dynamiques, c'est qu'elles véhiculent les caractéristiques d'une relation orientée directement vers des objets réels.²³ Les interfaces graphiques de manipulation directe ne sont cependant tangibles que si la manipulation n'est ni trop rudimentaire, ni trop indirecte (Beaudoin Lafon 1997). La conception des interfaces graphiques comme groupant sur l'écran des objets manipulables conduit donc soit à élargir le caractère tangible de l'objet pour y inclure les propriétés spatiales des éléments graphiques soit à concevoir la dimension physique de la relation à ces propriétés quitte à simplifier la manipulation (Fishkin et al., 2000).

CONCLUSION

Les pensées qui se forment dans le cours d'une action sont orientées sur les objets physiques que nous manipulons et elles sont activées par des environnements à la fois structurés et stabilisés selon des formats informationnels divers (Hammond et al., 1995 ; Kirsh, 1995). Une façon ordinaire de stabiliser cet environnement est de le peupler d'objets familiers interdépendants (couteau/fourchette/assiette). Comme le dit Norman, nos pensées sont « récupérées » en rencontrant dans l'espace de travail les objets qui les représentent²⁴. Les objets usuels pourvus d'*affordances* deviennent les supports fidèles de nos mouvements routinisés.

²² Je suis ici le constat fait par les philosophes de la perception selon lequel il existe des contenus non conceptuels (spatiaux et visuels) et que ces contenus non conceptuels interviennent de façon privilégiée dans la régulation de l'action.

²³ Corazza et Dokic (1993) soulignent que le caractère non objectif des représentations indexicales peut conduire soit à nier qu'elles possèdent un contenu descriptif car elles sont dépendantes de la vision et des *relations existentielles et expérientielles* avec des objets réels, soit à leur concéder un contenu descriptif propre qui serait un contenu circonstanciel. La seconde position attribuée à Perry conduit à dire qu'il existe des pensées indexicales ou des pensées égocentrées qui ont un contenu descriptif différent des représentations objectives.

²⁴ Voir supra note 2.

Les deux exemples mentionnés montrent qu'il existe un continuum entre les objets quotidiens et les artefacts informationnels pour deux raisons. D'abord ce sont les mêmes mécanismes de couplage qui assurent le contrôle de l'action. Les interfaces des billetteries comme les gobelets *Starbucks* ne sont pas principalement utilisés comme des instruments de communication de contenus conceptuels, mais d'abord comme des instruments d'interaction ou de support à l'action. Ensuite plus un objet est tangible plus il est efficace comme porteur d'informations pour la manipulation. Les objets manipulables, disposés dans l'espace sur une surface, rendent manifeste la prédominance des informations spatiales et visuelles même lorsqu'ils affichent des indications pour l'action car toutes les inscriptions graphiques placées sur un affichage possèdent des propriétés spatiales de contrôle.

De nombreux observateurs aujourd'hui proposent d'étendre les principes des interfaces de manipulation directe en utilisant les propriétés interactives des objets (Ishii, 2004 ; Beaudoin Lafon, 1997)²⁵. Ils nous rappellent qu'une partie de la cognition humaine s'ancre dans la manipulation directe des objets quotidiens. Ces principes s'appuient de façon explicite sur un modèle pragmatique de l'interprétation de l'information qui s'éloigne de l'idéal d'une communication homme-machine.²⁶ Jusqu'où ce modèle peut-il être poussé ? Quel est le lien entre ce modèle de l'action et la forme de l'objet ?

Dans ce modèle centré sur la manipulation des objets, l'exécution (ou l'accomplissement de l'action) reprend son autonomie car c'est en ligne que se connectent décision et exécution. Dans nos deux exemples, la détection de signes codés sur une surface manipulable vise à concevoir l'interaction avec l'information comme une interaction physique instrumentale. C'est parce que les inscriptions graphiques possèdent des propriétés spatiales et parce que le support spatial est immédiatement manipulable que l'objet s'avère susceptible de contrôler les mouvements corporels.

Bibliographie

- Agre, P. (1997). *Computation and Human Experience*, Cambridge, Cambridge University Press.
- Ballard, D., Hayhoe, M., Pook, P. & Rao, R. (1997). Deictic codes for the embodiment of cognition, *Behavioral and Brain Sciences*, n° 20 (vol. 4), pp. 723-67.
- Beaudoin Lafon, M. (1997). Interaction instrumentale : de la manipulation directe à la réalité augmentée, *Journées sur l'Interaction Homme-Machine*, Poitiers, Cepaduès Editions.
- Byrne, R. & Russon, A. (1998). Learning by imitation : a hierarchical approach, *Behavioral and Brain Sciences*, n° 21, vol.1, pp. 667-721.
- Clark, A. (1995). « Moving Minds », *Philosophical Papers*, Vol. 9, pp. 89-104.
- Clark, A. (1997). *Being There*, Cambridge, MA, MIT Press.
- Conein, B. & Jacopin, E. (1994). Action située et cognition : le savoir en place, *Sociologie du Travail, Travail et Cognition*, n° 4, vol. 36, pp. 475-500.
- Conein, B. (1997). L'action avec les objets : un autre visage de l'action située, *Cognition, Information et Société*, Raisons Pratiques, n° 8, sous la direction de B. Conein, & L. Thévenot, Editions de l'Ecole des Hautes Etudes, pp. 25-46.

²⁵ Je remercie Marc Relieu pour la référence d'Ishii, et Rémi Dorat pour celle concernant Beaudoin Lafon.

²⁶ L'apparition à Xerox Parc et leur diffusion par Apple des interfaces de manipulation directe constitue un événement majeur de l'histoire des technologies cognitives informatisées.

- Conein, B. & Jacopin, E. (1999). L'appauvrissement de la représentation : projection de plan et contrôle de l'action par les objets, sous la direction de C. Lenay & V. Havelange, *Mémoire de la technique et technique de la mémoire*, Technologie, Idéologie, Pratique, Erès, pp. 129-40.
- Corazza, E. & Dokic, J. (1993). *Penser en contexte*, Paris, L'Éclat.
- Dokic, J. (1999). L'action située et le principe de Ramsey, *La logique des situations*, Collection Raisons Pratiques, Vol. 10, Editions de l'EHESS, pp. 131-56.
- Fishkin, K., Gujar, A., Harrison, B., Moran, T. & Want, R. (2000). Embodied User Interfaces for Really Direct Manipulation, *Communications of the ACM*, Vol. 43, n° 9, pp.75-80.
- Gibson, J. (1979). *The ecological approach to visual perception*, Hillsdale, NJ, Lawrence Erlbaum.
- Goody, J. (1993-1994). *The interface between the written and the oral*, Cambridge, C.U.P. ; tr. fr. *Entre l'oralité et l'écriture*, Paris, P.U.F.
- Hammond, P., Converse, T. & Gras, J. (1995). The Stabilization of Environments, *Artificial Intelligence, Computational research on interaction and agency*, Vol. 72, n° 1-2, pp. 305-28.
- Hollan, J., Hutchins, E., Kirsh, D. (2000). Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research, *ACM Transactions on Computer-Human Interaction*, Vol. 7, No. 2, June, pp. 174-96.
- Hutchins, E. (1995). *Cognition in the Wild*, Cambridge, MIT Press.
- Ishii, H. (2004). Bottles : a transparent interface as a tribute to Mark Wieser, *IEICE, Transactions on Information Systems*, Vol. 87, n°6, pp. 1299-311.
- Kirsh, D. (1995). The Intelligent Use of Space, *Artificial Intelligence, Computational research on interaction and agency*, Vol. 72, n° 1-2, pp. 31-68.
- Kirsh, D. (1999). Distributed Cognition, Coordination and Environment Design, *Proceedings of the European conference on Cognitive Science*, pp. 1-11.
- Kirsh, D. (2001). The context of work, in *Human Computer Interaction*, 16, (2), pp. 306-322.
- Livet, P. (2000). La perception de l'action, *De la perception à l'action*, Paris, Vrin, pp. 219-236.
- Marr, D. (1982). *Vision*, Cambridge, MIT Press.
- Norman, D. (1988). *The psychology of everyday things*, New York, Basic Books.
- Norman, D. (1991). Cognitive artefacts, in Carroll J. (ed.) *Designing Interaction: Psychology at the Human-Computer Interface*, Cambridge, Cambridge University Press, tr. fr. 1993, collection *Raisons Pratiques*, Vol. 4, *Les objets dans l'action*, Editions de l'EHESS.
- Pacherie, E. (2000). Les différents niveaux de contenus perceptuels, *De la perception à l'action*, Paris, Vrin, pp.59-82.
- Suchman, L. (1987). *Plans and Situated Actions: the problem of human-machine interaction*, Cambridge, C.U.P.
- Tomasello, M. (1999). *The cultural origins of cognition*, Cambridge, Harvard University Press.
- Ullman, S. (1996). *High Level Vision*, Cambridge, Mass., MIT Press.