

Modèles hiérarchiques des représentations d'actions : apports de la psychopathologie et de l'imagerie par résonance magnétique fonctionnelle

Par Valérian CHAMBON

Thèse de Doctorat de Sciences Cognitives
mention Neurosciences

sous la direction de Nicolas FRANCK Chloé FARRER
présentée et soutenue publiquement le 10 novembre 2009

Membre du jury : Nicolas FRANCK, Professeur des Univ - Praticien Hosp., Université Lyon 1
Anne GIERSCH, Chargée de recherche, Université de Strasbourg Patrick HAGGARD, Professeur
d'université, University College London Chloé FARRER, Chargée de recherche, Université Toulouse 3
Etienne KOEHLIN, Directeur de recherche, INSERM Elisabeth PACHERIE, Directrice de recherche,
École Normale Supérieure de Paris

Table des matières

Contrat de diffusion . . .	6
Introduction Générale . . .	7
Partie 1 L'architecture du contrôle cognitif dans la schizophrénie Un modèle hiérarchique des représentations d'actions dans le cortex préfrontal dorsolatéral . . .	8
Introduction . . .	8
1 Des Fonctions exécutives au contrôle cognitif . . .	9
1.1 Fonctions Executives : nouveauté et contrôle . . .	9
1.2 Le syndrome dysexécutif : un marqueur de vulnérabilité de la schizophrénie ? . . .	10
1.3 Des outils pour évaluer les fonctions exécutives (planification, abstraction, attention et flexibilité cognitive) . . .	10
2 Profil exécutif et symptômes cliniques : une ou des schizophrénies ? . . .	12
2.1 Les outils en question : « Can the neuropsychological case-study approach be applied to schizophrenia ? » ¹ . . .	13
2.2 Troubles exécutifs et mécanismes sous-jacents : des fonctions exécutives aux processus de contrôle . . .	14
3 L'architecture du contrôle cognitif . . .	14
3.1 Un modèle <i>en cascade</i> du contrôle cognitif . . .	15
3.2 Les prédictions du modèle . . .	18
3.3 Présentation générale de l'étude . . .	18
The Architecture of Cognitive Control in Schizophrenia . . .	20
4. Discussion (1) . . .	29
Organization of Cognitive Control within the Lateral Prefrontal Cortex in Schizophrenia . . .	30
5. Discussion (2) . . .	40
5.1 Hypo-activation préfrontale et dysconnectivité rostro-caudale . . .	40
5.2 Du contrôle de l'action à sa compréhension . . .	43
Partie 2 Les mécanismes d'attribution d'intentions chez le sujet sain et le patient schizophrène Un modèle hiérarchique des représentations d'actions intentionnelles en 3ème personne . . .	45
Introduction . . .	45
1 Mindreading : comprendre et prédire le comportement. . .	46
1.1 De la théorie-théorie aux processus de simulation motrice. . .	46
1.2 La variété des intentions . . .	49
2. Les capacités de représentation et d'attribution d'intentions dans la schizophrénie . . .	52
3 La reconnaissance d'intentions : un cas particulier d'inférence probabiliste ? . . .	54
3.1 Vraisemblance perceptive et informations <i>a priori</i> . . .	55
3.2 L'apport du formalisme Bayésien . . .	56
4 Protocoles expérimentaux : intention motrice, intention superordonnée, intentions sociales . . .	59
Facteur information visuelle (pré-test) . . .	60
Facteur probabilité . . .	61

Etude chez le sujet sain - What are they up to? The role of sensory evidence and prior knowledge in action understanding . . .	61
1. Introduction . . .	62
2. Methods and results . . .	65
3. General discussion . . .	86
4. Conclusion . . .	92
Acknowledgments . . .	92
Appendix . . .	93
References . . .	96
- Etude chez le patient schizophrène - Mindreading under influence: abnormal dependence on prior knowledge in patients with schizophrenia . . .	100
1. Introduction . . .	100
2. Method . . .	103
3. Design and statistical analyses . . .	108
3. Results . . .	110
5. Discussion . . .	118
6. Conclusion . . .	121
Acknowledgments . . .	121
References . . .	122
5. Discussion générale . . .	124
5.1 Le rôle de l'information perceptive et des attentes <i>a priori</i> dans la compréhension de l'action . . .	124
5.2 Mentaliser sous influence : le cas de la schizophrénie . . .	126
5.3 Un modèle hiérarchique des représentations d'actions intentionnelles : système-miroir et <i>predictive coding</i> . . .	128
5.4 Du système-miroir au réseau de la Théorie de l'Esprit (<i>ToM network</i>) : les bases neurales du <i>mindreading</i> . . .	130
5.5 Etude IRMf : problématique générale . . .	132
Etude IRMf . . .	133
6 Comment le cerveau représente-t-il la variété des intentions ? . . .	133
6.1 Participants . . .	133
6.2 Description de la tâche (Figure 2) . . .	134
6.3 Analyses . . .	137
6.4 Résultats . . .	140
6.5 Discussion préliminaire – tâche SINGLE . . .	148
6.6 Conclusion préliminaire . . .	153
Reprise & perspectives . . .	155
Références bibliographiques . . .	159
– Partie I – L'architecture du contrôle cognitif dans la schizophrénie . . .	159
– Partie II – Les mécanismes d'attribution d'intentions chez le sujet sain et le patient schizophrène . . .	168
Publications . . .	179
Articles soumis ou en préparation . . .	179
Articles acceptés pour publication . . .	179

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas d'utilisation commerciale - pas de modification](https://creativecommons.org/licenses/by-nc-nd/2.0/fr/) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

Introduction Générale

L'objectif de ce travail est d'explorer deux des principaux troubles cognitifs que l'on recense dans la schizophrénie : le syndrome dysexécutif, ou trouble du contrôle de l'action volontaire, et les anomalies de représentation ou de compréhension du comportement d'autrui – qui prennent habituellement la forme d'une incapacité (ou d'une capacité réduite) à juger et à se représenter les états mentaux d'un tiers. Ces deux anomalies du fonctionnement cognitif ont en commun d'avoir un impact extrêmement délétère sur le fonctionnement quotidien des patients. Ils souffrent également d'être assez mal spécifiés, à la fois au plan de leur sévérité et de leur association, souvent diffuse, avec les différentes dimensions symptomatologiques de la maladie.

Nous proposerons dans les deux volets de cette thèse de réévaluer ces troubles à la lumière de deux modèles hiérarchiques des représentations d'actions – en première (Koechlin et *al.*, 2003) et en troisième personne (Kilner et *al.*, 2007). L'enjeu conjoint de ces modèles est de reconsidérer les fonctions de contrôle et de compréhension de l'action au niveau de leurs processus les plus élémentaires. Le modèle de Koechlin et collaborateurs propose à ce titre de décomposer les mécanismes de contrôle de l'action en trois niveaux distincts, fonction de la nature de l'information à contrôler (sensorielle, contextuelle, épisodique) et de l'implémentation de ces contrôles dans le cortex préfrontal dorsolatéral, le long d'un gradient de complexité antéropostérieur. Le modèle de Kilner, quant à lui, convoque le formalisme du *predictive coding* pour rendre compte de la robustesse de nos inférences mentales en situations perceptives d'ambiguïté ; il suggère, en particulier, que le cerveau est capable de suppléer cette absence d'information perceptive en se référant à des *attentes* constituées dynamiquement à chacun des niveaux de la hiérarchie corticale engagée durant l'observation.

Les tâches que nous présenterons en première et deuxième partie de ce travail adaptent, ou s'inspirent directement du formalisme de ces modèles. Ces modèles offrent selon nous un cadre cohérent, et biologiquement plausible, dans lequel explorer en détails la dynamique interactive des processus de contrôle et de compréhension de l'action, à la fois dans la mise en œuvre des comportements qu'ils gouvernent et des dysfonctions qu'ils peuvent occasionner.

Partie 1 L'architecture du contrôle cognitif dans la schizophrénie

Un modèle hiérarchique des représentations d'actions dans le cortex préfrontal dorsolatéral

Introduction

La schizophrénie représente aujourd'hui un objet d'investigation à part entière des sciences cognitives et les progrès enregistrés dans ce domaine sont inestimables – même s'il tarde encore à leur trouver des applications thérapeutiques ou réhabilitatives directes. Conséquence de cette ligne de recherche féconde, il est à présent largement admis que les patients schizophrènes souffrent de troubles variés du fonctionnement cognitif. Au plan du handicap, ces troubles ont une incidence dramatique sur la vie quotidienne des patients ; c'est particulièrement vrai des atteintes observées dans le registre des fonctions dites « exécutives », lesquelles sont connues pour sous-tendre la mise en œuvre de comportements élaborés. Or, la plupart de ces troubles dysexécutifs résistent à l'action médicamenteuse, aux traitements antipsychotiques en particulier dont l'efficacité sur les symptômes caractéristiques de la pathologie est pourtant reconnue. En outre, ces troubles présentent une variabilité exceptionnelle sur laquelle achoppent la plupart des outils et des modèles courants de compréhension : souvent développés sur la base d'observations cliniques et à des fins de détection ou de quantification du handicap, ces modèles demeurent en effet souvent muets quant aux liens qu'entretiennent déficits exécutifs et mécanismes neurocognitifs sous-jacents. Or, l'atteinte de ces mécanismes pourrait contribuer, non seulement à l'émergence de ces déficits, mais également à celle des symptômes *constitutifs* de la maladie elle-même. Rater le détail de ces mécanismes pourrait donc conduire à manquer ce qui, dans la pathologie, fait la spécificité du syndrome dysexécutif.

Les deux études que nous présentons dans la première partie de ce travail n'ont pas été initialement conçues pour la pathologie, mais trouvent leur origine dans un travail qui vise à mieux spécifier l'organisation hiérarchique du cortex préfrontal dorsolatéral et des différents modules qui le caractérisent. Ce travail – inspiré d'un protocole conçu à l'origine par Etienne Koechlin et collaborateurs (2003) – représente néanmoins une alternative crédible aux modèles d'investigation traditionnels des fonctions dites « frontales ». Cette alternative consiste à se situer à un niveau d'analyse du syndrome dysexécutif qui autoriserait sa décomposition en régions *spécifiques* de dysfonctionnement. Dans la schizophrénie, cette « parcellisation » fonctionnelle du syndrome – par ailleurs validée par des observations conduites en neuroimagerie fonctionnelle – vise à mettre en évidence

i) les niveaux de fonctionnement exécutif *spécifiquement* atteints chez les patients, d'une part,

ii) les relations qu'entretiennent ces régions de fonctionnement avec des mécanismes cérébraux sous-jacents, d'autre part,

iii) enfin, l'implication de certaines anomalies cérébrales dans la genèse peut-être *conjointe* des troubles exécutifs et des symptômes caractéristiques de la schizophrénie.

Le paradigme que nous présentons dans les pages qui suivent tente précisément de satisfaire à ces exigences. Elles nous paraissent cruciales, tant au point de vue de l'efficacité diagnostique qu'au plan de l'élaboration scientifique de la maladie.

1 Des Fonctions exécutives au contrôle cognitif

1.1 Fonctions Executives : nouveauté et contrôle

Les « fonctions exécutives » désignent un groupe de fonctions cognitives de haut niveau mobilisées pour la réalisation de comportements complexes, volontaires, adaptés aux sollicitations de l'environnement et dirigés vers un but. Ces fonctions de « contrôle » de l'action – souvent synonymes de « fonctions frontales » – incluent des capacités variées d'initiation, de planification et d'anticipation, d'orientation des ressources attentionnelles, de prise de décision, de résolution de problème, etc. Sous leur apparente hétérogénéité, ces capacités exécutives répondent ainsi au même objectif, celui de faciliter l'adaptation de l'individu à des situations nouvelles, lorsque les comportements déjà acquis ne suffisent plus (Shallice, 1982 ; Palmer & Heaton, 2000 ; Meulemans, 2004).

Les fonctions exécutives paraissent donc sollicitées chaque fois qu'une tâche requiert la mise en œuvre de processus contrôlés. Dans la conception théorique originale du « système exécutif » – défendue entre autres par Shallice (1982, 1988) –, on retient habituellement deux critères pour l'intervention de ces fonctions de haut niveau (voir Rabbit, 1997) :

1. Le premier de ces critères est celui de la nouveauté : les fonctions exécutives sont nécessaires pour réaliser des tâches nouvelles qui requièrent (a) la formulation d'un but, (b) la planification et le choix des différentes séquences comportementales qui permettront d'atteindre ce but, (c) la comparaison de ces plans entre eux, fonction de leur efficacité dans l'accomplissement du but poursuivi, (d) la mise en œuvre du plan sélectionné jusqu'à son accomplissement final, et enfin (e) l'éventuelle révision de ce plan en cas d'échec.
2. Le second de ces critères est celui de l'allocation contrôlée des ressources attentionnelles. Les fonctions exécutives seraient nécessaires pour initier de nouvelles séquences de comportement tout en interrompant des séquences en cours, ou en supprimant des réponses habituelles (routines). En d'autres termes, ces fonctions de haut niveau contrôleraient l'allocation des ressources attentionnelles chaque fois que l'environnement – ou les nouvelles exigences de la tâche en cours – commanderait le passage (switching) d'une séquence de comportement à une autre. Ce contrôle actif de l'attention, effectué sur de longues périodes de temps, favoriserait ainsi le déroulement de longues séquences de comportement. De fait, les fonctions exécutives seraient également indispensables pour contraindre la production de réponses inappropriées, pour coordonner la réalisation simultanée de deux tâches ainsi que pour détecter et corriger des erreurs (dans le programme moteur en cours, par exemple).

1.2 Le syndrome dysexécutif : un marqueur de vulnérabilité de la schizophrénie ?

Une altération des fonctions exécutives fait partie du tableau de la schizophrénie. Il est en effet admis, depuis les observations initiales de Bleuler (1911) et Kraepelin (1919), que les patients schizophrènes présentent de nombreuses difficultés pour organiser leurs actions. Celles-ci peuvent se traduire par des troubles de la planification du quotidien ou, de manière plus spécifique, par une incapacité à contrôler ses actions et les représentations qui lui sont associées (Frith, 1992).

Parce qu'ils se manifestent dans l'exercice d'une gamme variée de compétences souvent cruciales pour la conduite autonome de l'individu et pour la mise en œuvre de comportements adaptés, ces déficits ont un important retentissement sur la vie quotidienne du patient et en particulier sur sa vie professionnelle (Lysaker et al., 1997 ; Peuskens et al., 2005 ; Kurtz et al., 2005). S'ils ne peuvent expliquer toute la symptomatologie schizophrénique, ils sont néanmoins très fréquents dans cette affection (Sharma & Harvey, 2000) et sont reconnus pour affecter indifféremment les patients traités et non traités (Mortimer, 1996), chroniques et non chroniques (Hoff et al., 1992). Enfin, en raison de la fréquence et de l'étendue de leurs manifestations, il a été suggéré que ces troubles du fonctionnement exécutif pouvaient constituer un marqueur de vulnérabilité, voire un marqueur de trait de la schizophrénie (Zalla et al., 2004 ; Bozikas et al., 2006).

Il est aujourd'hui largement établi que le cortex préfrontal joue un rôle central dans l'exercice de ces fonctions cognitives de haut niveau que sont les fonctions exécutives (Miller & Cohen, 2001 ; Eisenberg & Berman, 2009, pour revue). Plusieurs études ont en effet montré que des lésions diffuses ou focales de cette région cérébrale affectaient considérablement les compétences exécutives de planification, de mémoire de travail et de prise de décision, ainsi que l'organisation temporelle des comportements volontaires (Shallice, 1988 ; Fuster, 1997). De fait, les troubles du fonctionnement exécutif observés dans la schizophrénie sont très souvent associés à des diminutions de l'activité préfrontale, en particulier dans sa région dorsolatérale (SPECT : Weinberger & Berman, 1996 ; IRMf ; Perlstein et al., 2003 ; Barch et al., 2001 ; MacDonald et al., 2005 ; Snitz et al., 2005). Cette « hypofrontalité » serait par ailleurs plus marquée chez le patient lorsque celui-ci doit réaliser des tâches requérant la mise en œuvre de compétences exécutives spécifiques (*test du Wisconsin* : Volz et al., 1997 ; *Tour de Londres* : Andreasen et al., 1992 ; *Stroop* : Carter et al., 1997). Dans plusieurs études conduites en neuroimagerie structurale, des altérations du volume de matière grise ont également été rapportées dans des régions sous-corticales fonctionnellement et anatomiquement reliées au cortex préfrontal (amygdale, hippocampe, striatum et ganglions de la base ; Velakoulis et al., 2006 ; Mamah et al., 2007 ; Wang et al., 2008). Selon certains auteurs, une communication inappropriée ou altérée entre ces régions sous-corticales et le cortex préfrontal pourrait rendre compte, plus spécifiquement, des troubles exécutifs dont souffrent conjointement les patients schizophrènes et certains patients frontolésés (Goldman-Rakic & Selemon, 1997). Le syndrome dysexécutif observé dans la pathologie ne serait donc pas explicable par un simple modèle de lésion focale du lobe frontal (Weinberger & Berman, 1996), mais exigerait de fait la contribution d'un réseau cérébral beaucoup plus large (Collette et al., 2005 ; Wang & Deutch, 2007).

1.3 Des outils pour évaluer les fonctions exécutives (planification, abstraction, attention et flexibilité cognitive)

L'origine historique des « fonctions exécutives » coïncide avec les premières tentatives conduites en neuropsychologie et en neurologie afin de circonscrire les fonctions du cortex préfrontal (Stuss & Benson, 1986, pour revue). Une grande variété de tests neuropsychologiques a donc été élaborée afin d'évaluer ces fonctions de haut niveau, tests qui ont fait leurs preuves dans de nombreuses pathologies neurologiques et psychiatriques, dont les traumatismes crâniens (Mattson & Levin, 1990), les maladies dégénératives (Dalrymple-Alford et al., 1994) et la schizophrénie (Mahurin et al., 1998). Les schizophrènes ont en effet des performances souvent comparables à celles des patients frontaux, en particulier dans les tâches requérant l'application de stratégies adaptées, la planification et la sélection de buts (tâches de la Tour ; Morice & Delahunty, 1996), la détection de règles et la formation de concepts (test du Wisconsin ; Laws, 1999, pour revue) ou la mise en œuvre de processus attentionnels de haut niveau (test de Stroop ; Hepp et al., 1996).

La *spécificité* des déficits évalués par ces batteries de tests ne fait néanmoins pas consensus, pour des raisons théoriques et méthodologiques sur lesquelles nous reviendrons plus loin. Pour autant, la *sensibilité* de ces tests aux troubles des fonctions exécutives – dans la schizophrénie comme dans toutes les pathologies où elles sont atteintes – a depuis longtemps fait ses preuves :

· **Tâches de la Tour** (Shallice, 1982 ; Burgess & Shallice, 1996a, b) _

Il existe plusieurs versions de la tâche de la Tour, mais elles sont toutes gouvernées par le même principe : déplacer un ensemble d'objet (des anneaux ou des perles) d'une position initiale à une position finale, et cela, en effectuant le moins de déplacements possibles. Cette tâche requiert de la part du sujet qu'il planifie l'ordre des déplacements qu'il effectuera, en respectant certaines règles fixées à l'avance (ne déplacer qu'un objet à la fois, par exemple). Dans les tâches de la Tour de Londres ou de la Tour de Hanoï, les patients schizophrènes réalisent un nombre significativement plus élevé de déplacements que les sujets contrôles (Morice & Delahunty, 1996). Ces difficultés pourraient refléter un déficit spécifique en planification, puisqu'il apparaît qu'elles ne sont pas imputables à un ralentissement moteur quelconque (Morris et al., 1995). Mais parce qu'elles requièrent de la part du sujet un ensemble de compétences variées (application de règles, inhibition de déplacements inappropriés, etc.), les tâches de la Tour pourraient toutefois ne constituer qu'une mesure aspécifique des capacités de planification (Marczewski et al., 2001).

· **Test de Stroop** (Stroop, 1935) _

Dans la version standard du test de Stroop, on présente aux sujets une série de mots, imprimés dans une couleur susceptible de ne pas correspondre à la couleur que ces mots désignent (condition interférente). La tâche consiste à nommer la couleur dans laquelle chaque mot a été imprimé (exercice qui requiert d'ignorer le contenu sémantique de ces mots). Parce qu'elle commande l'inhibition active d'une réponse automatique (une tendance à lire les mots plutôt qu'à nommer la couleur de ces mots), cette tâche a été considérée comme une mesure robuste des capacités d'attention sélective (ou « attention contrôlée »), et partant, du fonctionnement exécutif.

Dans la condition « interférente », de nombreuses études indiquent que les patients schizophrènes sont plus lents et font un plus grand nombre d'erreurs que les sujets sains (le patient ne peut s'empêcher de lire le mot écrit ; Hepp et al., 1996). Ces difficultés ont été interprétées dans les termes d'une sensibilité accrue des schizophrènes aux effets d'interférence couleur/contenu (et donc d'une perturbation des mécanismes inhibiteurs), puisque dans la condition « congruente » cette fois (contenu sémantique et couleur des

mots coïncident), les performances des patients sont davantage améliorées que celles des contrôles (Perlstein et *al.*, 1998, pour revue).

· **Le Wisconsin Card Sorting Test** (Grant & Berg, 1948)

Le test du Wisconsin est sans doute la mesure du fonctionnement exécutif la plus largement employée dans le champ de la psychopathologie (Goldman et *al.*, 1992). De manière très schématique, la tâche consiste à trier des cartes représentant des motifs simples, variant selon la forme, le nombre ou la couleur (« catégories »). La règle gouvernant le tri est choisie arbitrairement par l'expérimentateur, qui peut la modifier sans même en informer le sujet, dont la tâche consiste à induire cette règle par essais/erreurs successifs.

Dans le test du Wisconsin, les performances des patients schizophrènes révèlent des perturbations significatives en formation de concepts (subsumer les cartes présentées sous une même catégorie) et une flexibilité cognitive diminuée (erreurs de persévération en nombre élevé : les patients ont tendance à conserver la même stratégie lorsque celle-ci a fonctionné par ailleurs) (Van der Does & Van den Bosch, 1992, pour revue). Il a été suggéré que ces déficits pouvaient être associés, au moins partiellement, à des déficits plus fondamentaux en Mémoire de Travail (Goldman-Rakic, 1994), mais cette hypothèse fait encore débat (Stratta et *al.*, 1997a ; Hartman et *al.*, 2005).

2 Profil exécutif et symptômes cliniques : une ou des schizophrénies ?

Si l'évaluation clinique des troubles exécutifs dont souffrent les patients schizophrènes fournit le plus souvent des résultats robustes, il faut souligner néanmoins le caractère diffus et l'extrême variabilité des patterns de dysfonctionnement rapportés (Shallice et *al.*, 1991, pour revue). Cette variabilité, inter- et intra-individuelle, s'observe tant du point de vue du *nombre* de fonctions déficitaires que du point de vue de la *sévérité* des déficits. Bien plus importante chez les patients schizophrènes que chez les sujets contrôles, l'hétérogénéité de ces déficits pose la question de leur spécificité quant à la schizophrénie elle-même (O'Leary et *al.*, 2000).

Au sein du pattern global de perturbations cognitives dont souffrent les patients schizophrènes, les troubles exécutifs pourraient en effet constituer des déficits primaires, *spécifiques* de la maladie (Saykin et *al.*, 1991). Cette hypothèse, néanmoins, ne fait pas consensus, et selon certains auteurs le profil cognitif des patients schizophrènes traduirait plutôt l'existence d'une perturbation généralisée, excessivement diffuse, au cœur de laquelle émergeraient parfois des déficits plus saillants (exécutifs, par exemple), difficiles à systématiser, et par conséquent non spécifiques de la maladie elle-même (Heinrichs & Zakzanis, 1998).

Ce débat, encore très vif (Ragland, 2003 ; Bozikas et *al.*, 2006, pour revue), semble se poursuivre aujourd'hui pour au moins deux raisons, théorique et méthodologique, qui tiennent à la nature même de l'objet discuté (les fonctions exécutives) et à la pertinence des outils développés pour l'évaluer.

2.1 Les outils en question : « Can the neuropsychological case-study approach be applied to schizophrenia ? »¹

Le caractère hétérogène des troubles exécutifs recensés dans la schizophrénie invite en premier lieu à questionner la spécificité des tâches choisies pour les évaluer. Or, il est aujourd'hui établi que la plupart des tests conventionnels ne constituent pas de « pures » mesures du fonctionnement exécutif (Miyake et al., 2000; Rabbit, 1997 ; Palmer & Heaton, 2000, pour revue). Le test du Wisconsin en particulier est une tâche de résolution complexe qui requiert probablement la mise en œuvre, non pas d'une seule fonction, mais de processus cognitifs multiples – peut-être parallèles –, souvent basiques, et dont l'évaluation pourrait manquer les dysfonctionnements possibles (Anderson et al., 1991 ; Mountain & Snow, 1993, pour revue).

Cette contamination de la mesure exécutive par des processus routiniers, non exécutifs, est la source majeure des difficultés méthodologiques associées à l'évaluation des fonctions exécutives. A l'instar de la plupart des fonctions cognitives, le fonctionnement exécutif est en effet un « construit » théorique qui, par définition, ne peut jamais être *directement* mesuré. Comme le soulignent à juste titre Green et Nuechterlein (1999), les tests neuropsychologiques ne sont jamais que des « indicateurs » de ce construit, et, par conséquent, n'en réalisent que des mesures approximatives. C'est pourquoi l'évaluation d'une fonction (la planification, par exemple) requiert le plus souvent des mesures croisées (tâche des Tours, Wisconsin...). Or, cette multiplication des indicateurs présente un risque évident : celui d'accroître le nombre de processus interférents qui entrent dans la composition des mesures réalisées.

Non contrôlés et plus ou moins saillants selon les tests, l'existence tacite de ces processus pourrait donc expliquer la variabilité parfois importante des performances d'un sujet dans des tâches qui, pour être distinctes, évaluent néanmoins la même fonction (variabilité intra-individuelle). L'influence de ces processus interférents pourrait également expliquer que, dans une tâche exécutive quelconque, des comportements à première vue similaires puissent avoir des causes différentes (variabilité interindividuelle). Aux test du Wisconsin et de la Tour de Londres, l'hétérogénéité frappante des performances des patients schizophrènes pourrait ainsi traduire la contribution insidieuse de processus de second plan, non exécutifs et pour la mise en œuvre desquels les patients évalués ne seraient naturellement pas égaux (Anderson et al., 1991). De fait, la variabilité interindividuelle des schizophrènes en matière de planification traduirait en réalité l'extrême variabilité de leurs compétences à d'autres plans de la cognition (Van der Does & Van den Bosch, 1992).

Parce qu'elles mobilisent simultanément plusieurs fonctions cognitives, la plupart des tâches conventionnelles sont donc souvent *multidéterminées* (par des processus centraux et périphériques). Le problème méthodologique que pose la multidétermination des batteries de tests est, de fait, étroitement lié au problème théorique de la *multidimensionnalité* des fonctions exécutives. Les fonctions dites « exécutives » représentent en effet une notion théorique complexe, aux contours un peu flous, et dont les composantes ou dimensions se chevauchent partiellement. Elles sont à ce titre particulièrement difficiles à opérationnaliser et, en conséquence, les troubles qui les affectent s'avèrent difficiles à évaluer (Van der Linden et al., 2000, pour revue).

¹ Shallice, Burgess & Frith, 1991.

2.2 Troubles exécutifs et mécanismes sous-jacents : des fonctions exécutives aux processus de contrôle

La démarche qui caractérise les batteries de tests conventionnels présente un intérêt et une validité écologiques pourtant indiscutables – en matière de détection, de quantification ou de prise en charge du handicap. L'« impureté » caractéristique de ces tests, cependant, laisse souvent indéterminé le détail des mécanismes neurocognitifs spécifiquement atteints (Miyake et *al.*, 2000 ; Shallice et *al.*, 1991). La raison en est que ces batteries de tests ont été principalement élaborées pour détecter ou inférer la présence de lésions cérébrales, circonscrites (agnosies, apraxies...) ou diffuses (traumatismes crâniens, anoxies, démences d'Alzheimer...), et, à quelques rares exceptions près, n'ont pas été pensées en référence à un modèle cognitif particulier (Marzcewski et *al.*, 2001). Cette absence de référence théorique les rend par conséquent muettes quant aux liens qu'entretiennent troubles et mécanismes sous-jacents. Or, dans le champ de la psychopathologie en particulier, cette approche « descriptive » du déficit doit pouvoir céder le pas à une approche « compréhensive » de l'atteinte cognitive, chargée de mettre en évidence les mécanismes qui entretiennent ou contribuent à l'émergence des troubles étudiés. Selon un modèle influent de la schizophrénie, ces mécanismes pourraient en effet jouer un rôle important, non seulement dans la production des troubles cognitifs de la maladie, mais également *dans la production des symptômes psychotiques eux-mêmes* (Frith, 1992).

Dans la schizophrénie, les troubles du fonctionnement exécutif ne seraient donc pas assimilables à des atteintes « pures », clairement circonscrites sur le plan de la cognition, mais seraient étroitement liés aux dimensions de la maladie elle-même. De fait, il est depuis longtemps établi que ces dysfonctionnements sont très souvent associés à l'expression des symptômes négatifs de la schizophrénie (Hill et *al.*, 2001) ainsi qu'à ses symptômes de désorganisation (Mahurin et *al.*, 1998).

A ce titre, l'extrême variabilité des patterns de dysfonctionnement exécutif rapportés chez les patients schizophrènes pourrait refléter l'hétérogénéité clinique de la schizophrénie elle-même – *des schizophrénies ?* Or, si ces dysfonctionnements s'articulent en détail avec les expressions variées de la symptomatologie (productive, déficitaire, de désorganisation), alors il paraît indispensable d'outrepasser le plan de la pure phénoménologie et de mettre à jour, au-delà de la variabilité qui caractérise leurs manifestations, les mécanismes *communs* à ces dysfonctionnements. Cette entreprise, qui représente un enjeu crucial tant au niveau du diagnostic qu'au plan de l'élaboration scientifique de la maladie, requiert de se situer à un niveau d'analyse plus fin du déficit : en se déplaçant, par exemple, de la notion *reconstruite* de « fonction » (exécutive) à celle, moins théorique, de « processus » (de contrôle).

3 L'architecture du contrôle cognitif

Les fonctions dites « exécutives » sont, comme leur pluriel l'indique, des fonctions éminemment composites, produit de l'activité coordonnée de sous-processus cognitifs distincts (d'encodage, de sélection, d'inhibition, de manipulation de l'information, etc.). L'opération *concertée* de ces processus de traitement est cruciale pour la réalisation de comportements flexibles, adaptés aux exigences de l'environnement. Ces comportements s'inscrivent naturellement dans la durée, et, pour cette raison, requièrent le traitement

d'informations variées, parallèles et souvent asynchrones. Les processus chargés du traitement de ces informations se doivent par conséquent d'être coordonnés entre eux, afin de garantir l'unicité du comportement dans le temps de son exécution (cf. figure 1, p.19).

La coordination de ces différents sous-processus est assurée par un mécanisme de contrôle soutenu, appelé « contrôle cognitif » ou « contrôle exécutif »², dont la nature, la puissance, et l'implémentation ont donné lieu à un nombre considérable d'interprétations (Funahashi, 2001, pour revue). Au total, ce contrôle est reconnu pour sous-tendre une capacité cruciale, que nous mobilisons au quotidien : la capacité de coordonner nos pensées et nos actions en accord avec les buts qui les ont initiés.

Le caractère diffus des dysfonctionnements exécutifs recensés dans la schizophrénie pourrait en réalité traduire une altération plus profonde de ce mécanisme de contrôle – *commun* à toutes les fonctions exécutives –, et dont l'action souterraine gouvernerait la mise en œuvre de comportements élaborés. Or, si les dysfonctionnements exécutifs répertoriés dans la maladie sont relativement bien documentés, il reste encore à déterminer à quel(s) niveau(x) particulier(s) de ce mécanisme de contrôle s'inscrivent ces dysfonctionnements.

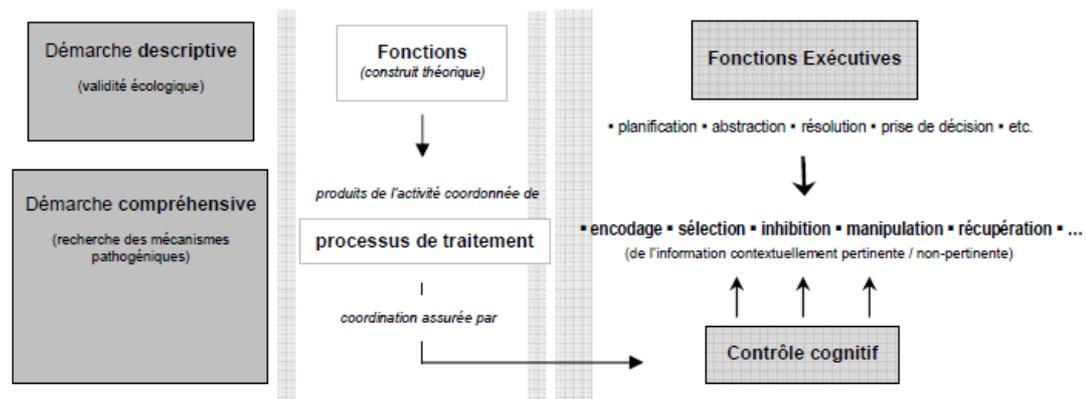


Figure 1. Des fonctions exécutives aux processus de contrôle (d'après Funahashi, 2001)

3.1 Un modèle en cascade du contrôle cognitif

En 2003, Etienne Koechlin et ses collaborateurs ont mis au point un protocole permettant d'évaluer dans le détail les capacités de sujets sains en matière de contrôle cognitif. L'intérêt de ce protocole tient en ce qu'il permet de distinguer *trois* niveaux distincts de contrôle. Chacun des ces niveaux est spécifié par la nature de l'information (sensorielle, contextuelle, épisodique) que le sujet doit *contrôler* pour accomplir la tâche avec succès. Ces trois types de contrôle sont associés à trois régions frontales distinctes (respectivement : prémotrice, caudale, et rostrale) dont l'activité co-varie avec le type de contrôle sollicité durant la tâche (Koechlin et al., 2003). Au sein du cortex préfrontal dorsolatéral humain, l'architecture du contrôle cognitif s'étagerait ainsi sur trois niveaux spécifiques de traitement :

1. un niveau de contrôle de l'information sensorielle (cortex prémoteur, BA 6) Ce niveau de contrôle est pris en charge par la région prémotrice de la hiérarchie frontale, connu pour être un répertoire d'associations Stimulus-Réponse (S-R). Le contrôle dit « sensoriel » est impliqué lorsque la sélection de la réponse motrice appropriée

² On dira à ce titre que les fonctions exécutives sont des fonctions placées « sous contrôle exécutif » (Funahashi, 2001).

- requiert de puiser dans ce répertoire l'association S-R appropriée au signal sensoriel qui survient.
2. un niveau de contrôle de l'information contextuelle (région caudale du cortex préfrontal, BA 9/44/45) Ce niveau de contrôle est pris en charge par la région caudale (postérieure) du cortex préfrontal, laquelle guide la sélection des représentations prémotrices (association S-R) élaborées au niveau sensoriel. Cette sélection est fonction des signaux contextuels qui accompagnent les stimuli sensoriels. Le contrôle contextuel assure ainsi la génération d'une réponse motrice adaptée au contexte dans lequel survient le signal sensoriel.
 3. un niveau de contrôle de l'information épisodique (région rostrale du cortex préfrontal, BA 10) Un événement qui a déjà eu lieu ou une information passée peuvent déterminer un épisode de traitement particulier, susceptible de moduler l'information de contexte et la réponse qui lui est habituellement associée. Dans la nomenclature de Koechlin et al. (2003), un épisode représente, non le contexte immédiat, mais le contexte d'une information passée. Cette distinction entre « contexte épisodique » et « contexte immédiat » fait écho à la distinction que suggèrent Park et ses collaborateurs (2003) entre contexte perceptif (caractères physiques/sensoriels du stimulus) et contexte cognitif (instructions ou informations pertinentes pour la tâche en cours, stockées en Mémoire de Travail). Dans une tâche expérimentale, par exemple, l'instruction qui précède un bloc inaugure un épisode de traitement particulier, qui détermine le ou les types de réponse que le sujet devra fournir pour accomplir la tâche avec succès. La sélection de l'instruction adéquate (qui déterminera la sélection de l'association Stimulus-Réponse pertinente) requiert donc un contrôle de nature épisodique ; ce type de contrôle sollicite, selon Koechlin et coll. (2003), la région rostrale du cortex préfrontal.

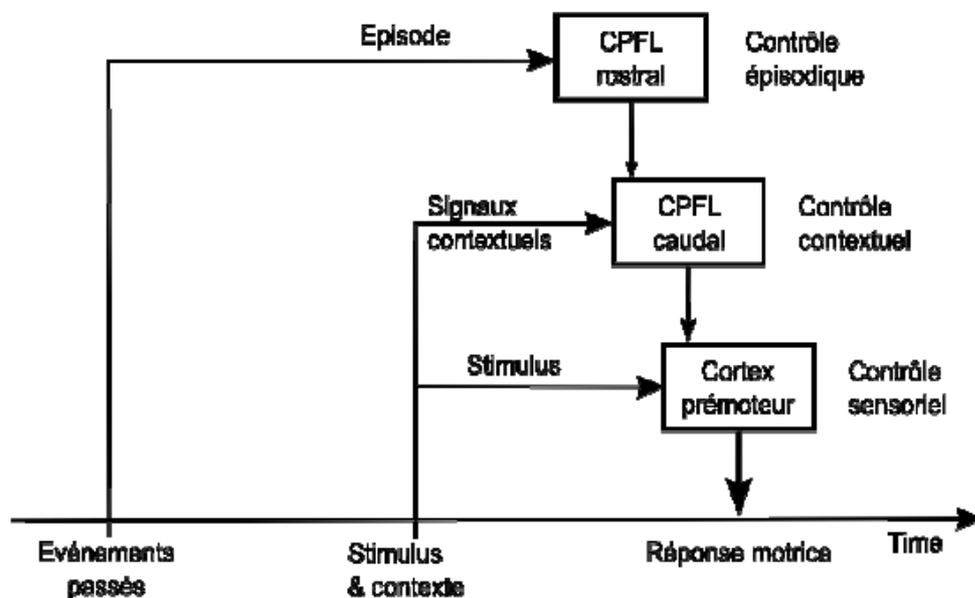


Figure 2. L'architecture en cascade du contrôle cognitif au sein du cortex préfrontal latéral (CPL). L'organisation multi-niveaux du traitement de l'information comprend un niveau de contrôle sensoriel impliqué dans la sélection des réponses motrices congruentes avec le signal sensoriel entrant (couleur du stimulus). Ce contrôle est pris en charge par les régions prémotrices latérales (aire de Brodmann, BA 6). Le niveau de contrôle contextuel

est impliqué dans la sélection des représentations prémotrices (i.e. des associations S-R) congruentes avec les signaux contextuels qui accompagnent les stimuli. Ce contrôle est pris en charge par la partie caudale du CPL (BA 9/44/45). Le niveau de contrôle épisodique est quant à lui impliqué dans la sélection des représentations caudales du CPL (i.e. tâches spécifiques ou ensembles d'associations S-R évoquées dans le même contexte) selon l'épisode temporel dans lequel les stimuli surviennent. Ce contrôle est pris en charge par la partie rostrale du CPL (BA 46/10). D'après Koechlin et al., 2003.

Cette superposition des niveaux de contrôle détermine une architecture « en cascade » du contrôle cognitif. Dans le modèle de Koechlin et coll. (2003), le cortex préfrontal – qui supporte la mise en œuvre de ces différents types de contrôle – est une structure hiérarchisée à plusieurs niveaux de traitement. Ce modèle généralise la théorie classique du contrôle de Baddeley et Hitch (1974), basée sur un système exécutif central contrôlant des « systèmes esclaves » de traitement et de stockage modalité-spécifiques. Ici, chaque niveau de la hiérarchie frontale maintient actives des représentations (associations Stimulus-Réponse, informations contextuelles, buts ou instructions) que contrôlent activement les niveaux de traitement supérieurs.

Dans cette architecture à plusieurs niveaux, la circulation de l'information le long de l'axe frontal antéropostérieur est gouvernée par l'organisation temporelle même de ces représentations, plutôt que par leur contenu ou complexité interne (Koechlin et al., 2003 ; Koechlin & Summerfield, 2007). L'organisation des comportements dans la durée, et le coût variable de leur élaboration, sont donc ici la conséquence de l'organisation hiérarchique des représentations d'actions le long du CPL : les régions les plus antérieures de la cascade sont recrutées lorsque l'action à produire nécessite l'intégration d'informations de plus en plus lointaines dans le temps (sensorielles > contextuelles > épisodiques) et cette demande accrue d'intégration temporelle se traduit, chez les participants, par une augmentation linéaire des temps de réaction (Fuster, 2004, pour revue).

La planification de comportements élaborés requiert ces contrôles *successifs*, qui garantissent la fidélité de nos actions aux buts que nous formulons en permanence et leur adaptation aux sollicitations variables de l'environnement. Naturellement, ces contrôles ne sont pas tous, toujours nécessaires ; leur exécution dépend du caractère familier de la situation en cours et de la nécessité d'outrepasser les routines d'action habituellement mises en œuvre. Admettons, par exemple, que je me trouve chez moi lorsque le téléphone sonne :

(A) Dans ce cas, l'élaboration du comportement le plus approprié nécessitera un effort de traitement réduit : j'associerai à l'information sensorielle perçue (la sonnerie du téléphone), la réponse motrice adéquate (répondre au téléphone). La sélection de cette réponse dans mon répertoire moteur est ici assurée par un contrôle de type *sensoriel*. L'information contextuelle, en ce cas, est neutre, puisque familière. Et parce qu'elle est familière, elle ne requiert pas d'être contrôlée. Lorsque je me trouve chez moi, j'ai en effet pour habitude de répondre au téléphone lorsque celui-ci sonne (routine d'action).

Mais admettons à présent que je ne sois pas chez moi, mais chez un ami. Le téléphone sonne :

(B) Dans cette situation, l'information contextuelle (je ne suis pas chez moi) *survient* sur l'information sensorielle. Elle exerce un contrôle sur ma réponse (« répondre au téléphone ») en l'inhibant, et de fait, assure la production d'un comportement adapté au contexte (ne pas répondre au téléphone lorsque je ne suis pas chez moi).

Enfin, je pourrais encore ne pas être chez moi, mais avoir reçu l'*instruction* de répondre au téléphone si celui-ci sonne :

(C) Dans ce troisième cas de figure, l'instruction détermine un *épisode* particulier dans lequel je suis, cette fois, autorisé à répondre au téléphone. Cette information épisodique (l'instruction) a pour effet de contrôler ma réponse en « court-circuitant » l'information contextuelle d'une part (« je ne suis pas chez moi ») et la réponse associée habituellement à ce contexte d'autre part (ne pas répondre au téléphone). La mise en œuvre d'un comportement de ce type sollicite bien un contrôle de nature épisodique, c'est-à-dire la sélection d'une réponse adaptée à l'épisode (l'instruction) dans lequel survient le signal sensoriel (la sonnerie du téléphone).

3.2 Les prédictions du modèle

L'intérêt majeur du protocole de Koechlin et coll. (2003) tient en ce qu'il propose des conditions expérimentales distinctes permettant de faire varier, de manière indépendante ou liée, la contribution de ces trois niveaux de contrôle cognitif (sensoriel, contextuel, épisodique). La mise en œuvre de l'un ou l'autre de ces contrôles a un coût variable, que le formalisme mathématique de la Théorie de l'Information (Shannon, 1948) permet de quantifier en valeurs objectives. Nous avons vu que le contrôle cognitif avait pour fonction d'assurer la sélection d'une réponse appropriée parmi un ensemble d'actions possibles, et cela, en vertu de signaux « de contrôle » de types potentiellement distincts (sensoriel et/ou contextuel et/ou épisodique). Or, il a été montré que le coût – ou la complexité – du contrôle cognitif co-variait avec le poids de l'information véhiculée par ces signaux de contrôle (Berlyne, 1957). La valeur de cette information est, dans la théorie de Shannon, une fonction logarithmique de la fréquence inverse avec laquelle une association Stimulus-Réponse est habituellement sélectionnée parmi un ensemble d'associations possibles de même type, pour un signal de contrôle donné³ (Koechlin et al., 2003). La complexité du contrôle sollicité dépend donc étroitement de la valeur de cette information. Dans le formalisme de la Théorie de l'Information, cette valeur est quantifiée en unités de mesure discrètes, et additionnables, les « bits » (*binary digits*). Le modèle en « cascade » de Koechlin et collaborateurs prédit qu'une sollicitation accrue des différents types de contrôle – *i.e.* une augmentation de leur valeur associée – devrait accroître les temps de réaction des participants et, dans le cas d'une perturbation localisée pour l'un ou l'autre niveau de la hiérarchie, les taux d'erreurs devraient augmenter en conséquence. L'expérience princeps réalisée par les auteurs sur des sujets sains a largement vérifié ces prédictions (Koechlin et al., 2003).

Une analyse détaillée de ces taux d'erreurs et de ces latences offre donc la possibilité d'évaluer dans le détail les performances des participants pour chacun des trois niveaux de contrôle sollicités ; partant, elle pourrait nous permettre d'identifier de manière précise le niveau de contrôle cognitif éventuellement déficitaire dans la schizophrénie. Or, malgré une littérature considérable sur les altérations des fonctions exécutives dans la pathologie, l'architecture du système de contrôle qui sous-tend la mise en œuvre de ces fonctions n'a jamais été évaluée.

3.3 Présentation générale de l'étude

Dans l'étude qui suit, nous avons adapté l'expérience de Koechlin et collaborateurs (2003) à un groupe de 24 patients schizophrènes et 24 participants non-psychiatriques, appariés en âge, sexe, latéralité et niveau d'éducation. La tâche proposée aux participants était une

³ Ainsi, une association S-R fréquemment mise en œuvre requerra un contrôle peu coûteux (sinon nul dans le cas des routines d'action, par exemple).

tâche d'association visuo-motrice de type 'Go/No-go', dans laquelle les participants devaient associer, selon certaines règles, des réponses latéralisées (bouton droit, bouton gauche d'un boîtier-réponse) aux stimuli qui leur étaient présentés (des disques dans la première tâche, des lettres dans la seconde). Les stimuli variaient sur plusieurs dimensions que les participants devaient impérativement prendre en compte dans la production de leur réponse. Dans la tâche « Disque », seule la couleur des disque variait (information sensorielle). Dans la tâche « Lettres », la couleur des lettres variait également (information contextuelle), ainsi que leur casse (Majuscule-Minuscule) et/ou leur type phonémique (Voyelle-Consonne). Chaque bloc était précédé d'une inscription détaillant explicitement le type d'association S-R à réaliser (propriété du stimulus > réponse gauche / réponse droite / ne rien répondre). Les instructions représentaient les informations épisodiques de la tâche ; elles spécifiaient, au début de chaque bloc, le ou les buts que le sujet allait devoir poursuivre durant l'épisode de traitement à venir (pour une description détaillée du protocole, voir Chambon et al., 2008, ci-dessous).

Nos prédictions étaient les suivantes. Si les troubles du fonctionnement exécutif observés dans la schizophrénie témoignent bien, en amont, d'une altération d'un ou de plusieurs niveaux du contrôle cognitif, alors nous devrions observer chez les patients une augmentation des taux d'erreurs (ERs) et des temps de réaction (TR) dans les conditions expérimentales modélisant les facteurs

i) « **sensoriel** » ? Ce niveau de contrôle pourrait être relativement préservé dans la schizophrénie, comme le suggèrent certaines études évaluant les performances des schizophrènes pour le traitement des informations sensorielles (tâche de discrimination de couleurs de stimuli ; Posada & Franck, 2002).

ii) « **contextuel** » ? De nombreuses études rapportent chez les schizophrènes l'existence d'un trouble majeur de l'exploitation du contexte. Cette perturbation est relativement bien documentée depuis les études princeps de Stroop sur la lecture chronométrée des noms de couleurs (1935). Dans cette tâche requérant l'inhibition active d'un contexte perceptif interférent (la couleur des mots est parfois différente de la couleur que ces mots désignent), les patients schizophrènes présentent en effet des latences et des taux d'erreurs plus importants que les sujets sains (Henik & Salo, 2004). Des résultats similaires ont été rapportés dans des tâches d'attention sélective, nécessitant à la fois l'inhibition active de distracteurs attentionnels et le traitement sélectif de cibles contextuellement pertinentes pour la tâche en cours (Cohen & Servan-Schreiber, 1992). La sélection des informations pertinentes et l'organisation globale de la réponse supposeraient de pouvoir se référer à une représentation cohérente du contexte que les patients schizophrènes ne parviendraient pas à élaborer, et dont l'absence pourrait expliquer, pour une large part, les troubles attentionnels maintes fois observés dans la schizophrénie.

iii) « **épisodique** » ? Selon Jones et collaborateurs (1991), les troubles cognitifs diffus dont souffrent les patients schizophrènes pourraient traduire un déficit central primaire pour le traitement et/ou l'intégration des informations passées. En outre, il a été suggéré que les difficultés des schizophrènes en matière de planification pouvaient ressortir d'une impossibilité plus fondamentale à mettre en œuvre les règles adéquates gouvernant la tâche en cours (Posada & Franck, 2002). Cette difficulté pourrait être associée à un défaut de représentation de l'information épisodique – lorsque celle-ci est définie par une règle, ou instruction, qui spécifie l'épisode de traitement approprié. Plus généralement, cette difficulté pour le traitement ou la restitution des informations passées pourrait faire écho aux perturbations de la mémoire dite « épisodique », déjà observées dans la schizophrénie (Heinrichs & Zakzanis, 1998, pour revue). L'un des processus-clé de cette

mémoire est la capacité à lier entre eux d'une part les propriétés physiques de l'événement vécu, d'autre part l'épisode temporel ou le lieu dans lequel cet événement a survécu (*memory binding*). Il a été montré, par exemple, que les patients schizophrènes avaient des performances nettement altérées dans les tâches manipulant des informations temporelles (Rizzo et al., 1996) ou des fréquences d'apparition d'événements (Gras-Vincendon et al., 1994). Des perturbations identiques sont également observées dans les tâches de « reality monitoring », où le patient doit déterminer la source (lui-même ou l'expérimentateur) d'une information qui lui est présentée (Mechelli et al., 2007 ; Allen et al., 2007).

Le déficit des patients schizophrènes pourrait être plus saillant pour l'un ou l'autre de ces niveaux de contrôle. L'atteinte du niveau contextuel, par exemple, déterminerait, selon de nombreux auteurs, un profil cognitif très caractéristique de la schizophrénie (Cohen & Servan-Schreiber, 1992 ; Schooler et al., 2008). Nous avons donc comparé, dans l'étude qui suit, les performances des schizophrènes pour chacun de ces trois niveaux de contrôle ; nous avons également tenté de déterminer si ces performances corrélaient avec la sévérité de l'une ou l'autre des dimensions symptomatologique de la pathologie : productive (SAPS ; Andreasen, 1983), déficitaire (SANS ; Andreasen, 1984), ou de désorganisation.

The Architecture of Cognitive Control in Schizophrenia

Valérian Chambon, Nicolas Franck, Etienne Koechlin, Eric Fakra, Gabriela Ciuperca, Jean-Michel Azorin & Chloé Farrer

(Brain)

doi:10.1093/brain/awn032

Brain (2008), 131, 962–970

The architecture of cognitive control in schizophrenia

Valérien Chambon,¹ Nicolas Franck,^{1,2} Etienne Koechlin,³ Eric Fakra,⁴ Gabriela Ciuperca,⁵ Jean-Michel Azorin⁴ and Chloé Farrer¹

¹Centre de Neurosciences Cognitives, Université de Lyon, CNRS, 69675 Lyon, ²Centre Hospitalier le Vinatier and Université Claude Bernard, 69675 Lyon, ³INSERM, Université Pierre et Marie Curie and Ecole Normale Supérieure, 75005 Paris, ⁴Service Hospitalo-Universitaire de Psychiatrie, CHU Sainte-Marguerite, 13009 Marseille and ⁵Equipe de Probabilités, Statistiques et Physique mathématique, Université Claude Bernard, CNRS, 69366 Lyon, France

Correspondence to: Valérien Chambon and Chloé Farrer, Centre de Neurosciences Cognitives, UMR 5229 CNRS, 67, bd Pinel, 69 675 BRON Cedex, France
E-mail: valerian.chambon@isc.cnrs.fr; farrer@isc.cnrs.fr

Executive dysfunctions have long been considered a common feature of schizophrenia. However, due to their extreme heterogeneity, it is not clear whether these impairments take place at a particular level of executive functioning or non-specifically affect various aspects of behavioural control. To answer this question, we used an experimental paradigm based upon a multistage model of prefrontal executive function. This model postulates that cognitive control is organized in three hierarchically ordered control processes, operating with respect to the perceptual context (sensory and contextual controls) or the temporal episode in which the person is acting (episodic control). Twenty-four patients with schizophrenia and 24 non-psychiatric controls participated in two distinct experiments designed to separately assess each of these three levels of control. The results indicate that both sensory and episodic dimensions of cognitive control were spared in schizophrenic patients, but that they showed great difficulty in contextual conditions, as the selection of the appropriate response among competitive ones required taking into account information related to perceptual context. Contextual control can be considered as a set of executive processes mediating the hierarchical organization of behaviour. Patients' deficit in cognitive control therefore reflects a specific problem in the hierarchical control of action, leading to the selection of inappropriate behavioural representations for ongoing action plans. We also showed that this impairment was a good predictor of disorganization syndrome scores, suggesting that these clinical manifestations might result from a deficit in the combination or selection of hierarchically organized action representations.

Keywords: cognitive control; cascade model; schizophrenia; context processing; disorganization syndrome

Abbreviations: EPs = error percentages; RTs = reaction times; SR = stimulus-response; WAIS = Wechsler Adult Intelligence Scale

Received July 11, 2007. Revised November 9, 2007. Accepted February 12, 2008. Advance Access publication March 3, 2008

Introduction

Executive dysfunctions have long been considered a common feature of schizophrenia (Velligan and Bow-Thomas, 1999). However, despite their pervasiveness, these high-level cognitive disorders have proved highly resistant to systematization. Particularly noticeable is their extreme variability among patients in terms of both severity and nature (Shallice *et al.*, 1991), raising the question of their specificity regarding schizophrenia itself (O'Leary *et al.*, 2000) as well as the specificity of the tasks developed to assess them (Lezak, 1993; Axelrod *et al.*, 1996).

One way to solve this ambiguity is to tackle the problem at a more fine-grained level of cognitive functioning. Executive functions are considered to be a product of

various processes (e.g. information selection, inhibition, maintenance, etc.), the coordination of which is assumed to be achieved by a mechanism called cognitive control (Funahashi, 2001). Our claim is that executive function deficits in schizophrenia could be accounted for by a specific impairment of this control mechanism.

Despite notable advances, the executive processes and their functional architecture remain poorly specified (Godefroy *et al.*, 1999). Recently, Koechlin *et al.* (2003) addressed this question by proposing an original model of cognitive control, based on an extensive investigation of prefrontal function and organization. By demonstrating that the frontal cortex is functionally organized as a *cascade* of control processes, the authors showed that cognitive

© The Author (2008). Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

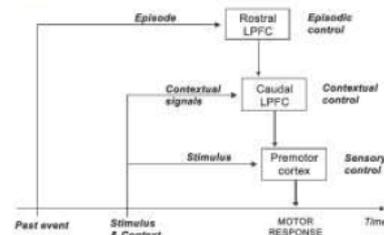


Fig. 1 Model of cognitive control by Koechlin et al. (2003). The multistage organization of information processing includes a SENSORY control level involved in selecting the motor responses that are the most appropriate to stimuli that occur. This control is subserved by the lateral premotor regions. The CONTEXTUAL control level is involved in selecting SR associations according to contextual signals accompanying stimuli occurrences. This control is subserved by the caudal part of the lateral prefrontal cortex. The EPISODIC control level is involved in selecting task-sets or consistent sets of SR associations evoked in the same context according to the temporal episode in which stimuli occur; that is, according to events that previously occurred or to ongoing internal goals. This control is subserved by the rostral part of the lateral prefrontal cortex. Courtesy of Koechlin et al. (2003).

control involves at least three levels of processing implemented in distinct frontal regions. These control processes operate with respect to the perceptual context (contextual control) or the temporal episode in which the person is acting (episodic control) (Koechlin et al., 2003). The model generalizes the classical theory of executive control from Baddeley and Hitch (1974), based on a central executive system controlling multiple maintenance slave systems, to a multistage architecture, where each level of the frontal hierarchy mediates the processing of a distinct signal (sensory, contextual or episodic signals) involved in controlling the selection of appropriate stimulus-response (SR) associations. Furthermore, the cascading nature of this model is derived from the idea that processing carried out at each level of this hierarchy is constantly informed by the processing driven by progressively higher levels, thus, giving rise to a cascade of top-down, successive controls (Fig. 1).

In addition to identifying different control processes associated with distinct types of signal, this model also takes into account variations in the demands of these controls. These demands vary as a function of the information (in terms of information theory) (Shannon, 1948) conveyed by the control signals that are required for selecting appropriate representations for action. The cascading model of Koechlin et al. (2003) thus proves appropriate, not only for identifying which processes of the control hierarchy might be specifically dysfunctional in schizophrenia, but also for evaluating the influence of the varying demands of these controls on patients' performance.

We adapted the experimental paradigm of Koechlin et al. (2003) in order to evaluate the performance of patients with schizophrenia, and healthy participants, in visuomotor association tasks modelling sensory, contextual and episodic controls. We expected impaired cognitive control to affect specific levels of the hierarchy. As it has been shown that patients correctly use a rule to associate a stimulus with a response (Posada and Franck, 2002), we predicted that the level of sensory control would be spared. In contrast, contextual and/or episodic controls could be more specifically impaired because they require higher temporal integration, a deficit which might characterize schizophrenia (Jones et al., 1991; Cohen and Servan-Schreiber, 1992; Gras-Vincendon et al., 1994). Based upon the many reports of an association between executive dysfunction and disorganized thoughts and behaviours in schizophrenia (Mahurin et al., 1998; Kravartii et al., 2005), we also looked at whether patients' performance in cognitive control was associated with the severity of disorganization symptoms.

Methods

Participants

Twenty-four patients with schizophrenia (9 females, 15 males; mean age: 37.33 years, SD: 9.7) and 24 healthy participants (9 females, 15 males, mean age: 36 years, SD: 12) participated in the study. Patients recruited fulfilled DSM-IV (American Psychiatric Association, 1994) criteria of schizophrenia, with no other psychiatric diagnosis on DSM-IV Axis I. All patients were receiving antipsychotic medication (principally olanzapine, risperidone and aripiprazole) and were clinically stable at the time of testing (duration of illness: mean: 9.4 years, SD: 6.3). Negative and positive symptoms were evaluated with the SANS (mean: 40.3, SD: 15.4) (Andreasen, 1983) and the SAPS (mean: 41.2, SD: 19) (Andreasen, 1984). A disorganization score (mean: 26.95, SD: 11.97) was also computed by summing the following subscores: bizarre behaviour, positive formal thought disorder, alolia and inappropriate affect. Intellectual efficiency was assessed in the schizophrenia group by two trained neuropsychologists using the Wechsler Adult Intelligence Scale (WAIS-third edition; Wechsler, 1997). The mean 'Total IQ' reported for the schizophrenia sample (mean: 92.2, SD: 8.57) did not differ from the normal range.

Controls subjects were matched with patients for sex, age, handedness (patients: mean = 0.74, SD = 0.51, controls: mean = 0.79, SD = 0.46; Oldfield, 1971) and years of education. None of them reported psychiatric problems.

For both groups, exclusion criteria included dyschromatopsy, history of neurological illness or trauma, alcohol or drug dependence according to DSM-IV criteria, analphabetism and age older than 60 years. All participants reported normal or corrected-to-normal visual acuity. After complete description of the study to the subjects, written informed consent was obtained. This research was approved by the local Ethical Committee.

Task

The study consisted of two behavioural experiments that were designed to separately vary the demands of sensory and episodic controls (MOTOR experiment) and contextual and episodic

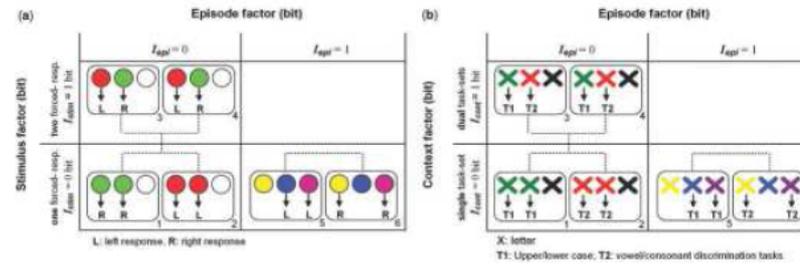


Fig. 2. Experimental designs. Rounded boxes represent behavioural episodes (numbered from 1 to 6) with related stimuli and instructions. (a) MOTOR experiment: episodes formed three distinct experimental conditions crossing the Episode factor with the Stimulus factor. In this experiment, stimuli were coloured discs. Subjects ignored distractor stimuli or responded by pressing the left (L) or right (R) response button. Dashed lines connect episodes involving congruent stimulus-response associations. (b) TASK experiment: episodes formed three distinct experimental conditions crossing the Episode factor with the Context factor. In that experiment, stimuli were letters (represented by the symbol X) and contextual signals were colours of letters. Letters were pseudorandomly chosen from the set {A, E, I, O, U, Y, a, e, i, o, u, y, B, D, G, K, R, T, b, d, g, k, r, t}. Depending on the contextual signals, subjects ignored letters (no arrow) or performed either a lower/upper case (T1) or a consonant/vowel (T2) discrimination task on letters. Dashed lines connect episodes involving congruent associations between contextual signals and task-sets.

controls (TASK experiment). For each experiment, participants performed a visuomotor association task in which they responded to a series of successively presented visual stimuli (coloured discs or letters) by pressing left or right hand-held response buttons. Each experiment was administered using a 6 × 6 Latin-square block design consisting of six different stimuli, each presented in six separate blocks (episodes). Each block included a series of 12 stimuli presentations (duration: 500 ms; onset asynchrony: 3500 ms), preceded by an instruction cue (episodic signal) lasting 3200 ms. Subjects were required to respond as quickly and accurately as possible.

The stimuli were presented on a computer monitor (16") and a head-fixation device was used to both control the distance separating the subject from the test monitor (60 cm) and to minimize head movements that might influence reaction times. The MOTOR and TASK experiments were conducted on two successive days. Their order of presentation was counterbalanced across participants, as was the order of block presentation within each experiment.

Prior to running the experiment, participants were trained in order to avoid possible biases due to learning effects during the test session. The stimuli were presented using the EXPE6 software (<http://www.ehess.fr/centres/lcp/expe>).

The demands of sensory, contextual and episodic control were separately varied across the experiments. These variations were quantified according to the computational model from Koechlin et al. (2003), based on Shannon's information theory (Shannon, 1948). Details concerning the calculation of information values for the different signals are provided in the original study by Koechlin et al. (2003).

In the MOTOR experiment, subjects had to respond to coloured discs by pressing the left (L) or right (R) response button and to ignore distractor stimuli. The demands of sensory control were conveyed by the stimulus indicating the response to be chosen and

varied across blocks. These demands were expressed in 'binary digits' (bits) and represented discrete values quantifying the amount of information subjects had to control during the task [noted (0) or (1)] (Koechlin et al., 2003). Sensory information was of 0 bit ($I_{sens} = 0$ bit) for blocks with one forced response and of 1 bit for blocks with two forced-responses ($I_{sens} = 1$ bit). The ratio of left versus right responses was equal to 1. In this experiment, no contextual signal was used, so that no contextual control occurred (Fig. 2a).

In the TASK experiment, subjects had to respond to coloured letters. Depending upon the colour of the letter, subjects had to perform a lower/upper case (T1), or a consonant/vowel (T2) discrimination task (using the left and right response buttons) or had to ignore the stimuli. The demand of contextual control was conveyed by the colour of the stimulus about a task-set and varied across blocks. Contextual information was of 0 bit for blocks involving a single task-set ($I_{cont} = 0$ bit) and of 1 bit for blocks involving a dual-task-set ($I_{cont} = 1$ bit). The ratio of trials associated with task-set T1 versus task-set T2 was equal to 1. Letters were pseudorandomly chosen so that in each block the ratio of left versus right responses and the ratio of congruent versus incongruent letters (same versus different responses for T1 and T2) were equal to 1. In that experiment, sensory control was maintained constant across the blocks (Fig. 2b).

For each experiment, the demand of episodic control (e.g. episodic information) was conveyed by instruction cues indicating which task-set to perform, given other signals (sensory or contextual signals). Episodic information was required for selecting appropriate SR associations (MOTOR experiment) or appropriate task-sets (TASK experiment) and varied parametrically across blocks (from $I_{epi} = 0$ to 1 bit) (Fig. 2). Proportions of two successive trials including identical stimuli or identical contextual signals were also maintained constant across blocks. In each block, sequences of stimuli were pseudorandomized so that the proportion of distractors was equal to 33%.

Description of the experimental blocks for experiment 1

Block #1: Discs were either green or white. White discs were distractors and subjects had to respond to green discs by pressing the right button (one forced-response episode, $I_{stim}[0]*I_{epi}[0]$).

Block #2: Discs were either red or white. White discs were distractors and subjects had to respond to red discs by pressing the left button (one forced-response episode, $I_{stim}[0]*I_{epi}[0]$).

Blocks #3 and #4: Discs were either green, red or white. Subjects had to respond to stimuli as in blocks #1 and #2 (two forced-response episodes, $I_{stim}[1]*I_{epi}[0]$).

Block #5: Discs were either yellow, blue or cyan. Yellow discs were distractors and subjects had to respond to blue and cyan discs by pressing the left button (one forced-response episode, $I_{stim}[0]*I_{epi}[1]$).

Block #6: Discs were either yellow, blue or cyan. Blue discs were distractors and subjects had to respond to yellow and cyan discs by pressing the right button (one forced-response episode, $I_{stim}[0]*I_{epi}[1]$).

Description of the experimental blocks for experiment 2

Block #1: Contextual signals were either green or white. White signals indicated subjects should ignore the letter. Green signals indicated subjects should perform task T1 (single task-set episode, $I_{cont}[0]*I_{epi}[0]$).

Block #2: Contextual signals were either red or white. White signals indicated subjects should ignore the letter. Red signals indicated subjects should perform task T2 (single task-set episode, $I_{cont}[0]*I_{epi}[0]$).

Blocks #3 and #4: Contextual signals were either green, red or white. Subjects had to respond to letters as in blocks #1 and #2 (dual task-set episode, $I_{cont}[1]*I_{epi}[0]$).

Block #5: Contextual signals were either yellow, blue or cyan. Yellow signals indicated subjects should ignore letters. Blue and cyan signals indicated subjects should perform task T2 (single task-set episode, $I_{cont}[0]*I_{epi}[1]$).

Block #6: Contextual signals were either yellow, blue or cyan. Blue signals indicated subjects should ignore letters. Yellow and cyan signals indicated subjects should perform task T1 (single task-set episode, $I_{cont}[0]*I_{epi}[1]$).

Data analyses

Reaction times (RTs) for correct responses and error percentages (EPs) were recorded and analysed using the software Statistica7. In each experiment, the six blocks formed three distinct experimental conditions crossing the episode factor (instruction cue) with either the Stimulus (MOTOR experiment) or the Context factor (TASK experiment). In the MOTOR experiment, the Stimulus factor (I_{stim}) contrasted one-forced response and two-forced responses episodes. In the TASK experiment, the Context factor (I_{cont}) contrasted single-task-set and dual-task-set episodes. Finally, in both experiments, the Episode factor (I_{epi}) contrasted episodes with $I_{epi}=0$ and episodes with $I_{epi}=1$. For the sake of clarity, the Episode factor was termed EpiMotor (I_{epiM}) in the MOTOR experiment and EpiTask (I_{epiT}) in the TASK experiment.

For each experiment, we computed a 2 (group) \times 2 (control factor 1) \times 2 (control factor 2) repeated-measures ANOVA on both RT and EP data. Analyses were made with group

(schizophrenic patients versus controls) as a between-subjects factor, Episode ($I_{epi}=0$ versus $I_{epi}=1$) as a within-subjects factor and Stimulus ($I_{stim}=0$ versus $I_{stim}=1$) and Context ($I_{cont}=0$ versus $I_{cont}=1$) as within-subjects factors for the MOTOR and TASK experiments, respectively. *Post hoc* Fisher tests were performed to identify differences. Whenever the variance structure did not conform to the requirements for parametric analyses, logarithmic transformations were used to obtain the required conformity.

Regression analyses

Regression analyses were conducted to evaluate the influence of patients' cognitive performance on their clinical scores. Independent analyses were made with the disorganization score, its subscores, the SANS and SAPS scores as dependent variables and cognitive performances as explanatory factors. Cognitive scores were computed by subtracting EPs in conditions in which $I=1$ from conditions where $I=0$, while maintaining the other factor constant. Three scores were computed: a sensory score ($Score_{sens}$), a contextual score ($Score_{cont}$) and an episodic score ($Score_{epi}$). For each clinical score, we conducted regression analyses using the different cognitive scores independently (simple linear regressions) or their transformed values (simple non-linear regressions with logarithmic, polynomial or exponential transformations) or a linear combination of two cognitive scores (multiple linear regressions). Models with the highest adjusted R -squared (R^2) and a P -value ≤ 0.05 are reported.

Results

Reaction times

ANOVAs performed on the transformed reaction times first confirmed the results obtained by Koechlin *et al.* (2003), showing significant effects of the Stimulus, Context and Episode factors [all effects: $F(1,138)$, $P < 0.0001$]. Reaction times were slower as the demands of cognitive controls increased (from $I=0$ to 1 bit). A group effect was also observed with patients being significantly slower than controls [both experiments: $F(1,138)$, $P < 0.0001$]. There were, however, no interaction effects between group and cognitive factors indicating that the generalized slowdown of patients' RTs was independent of the condition (Fig. 3).

Error percentages

Participants' error percentages were found to significantly increase as the Stimulus, Context and Episode factors varied from 0 to 1 bit [all effects: $F(1,138)$, $P < 0.0001$]. EPs were also higher in patients compared to controls, as revealed by a significant group effect [both experiments: $F(1,138)$, $P < 0.0001$]. Significant interactions between the Group factor and the Cognitive factors were only observed in the TASK experiment for both the Context and the Episodic factors [interactions: $F(1,138) = 11.97$, $P < 0.001$; $F(1,138) = 5.41$, $P < 0.05$]. Although patients performed better than chance (one-tailed t -tests, all $t < -25$, $df = 24$, $P < 0.001$), the percentage of errors they made was substantially higher as episodic and contextual information

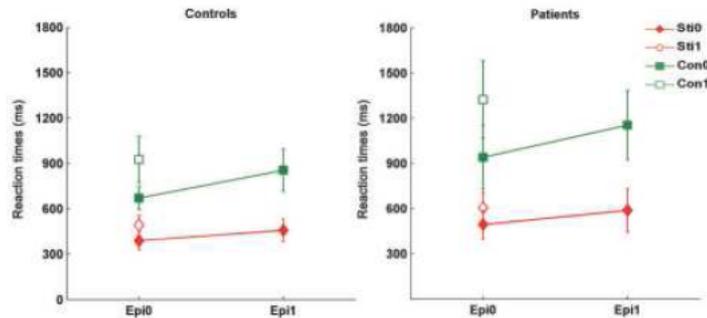


Fig. 3 Reaction times to stimuli (mean \pm SD patients and controls averaged across correct responses). Reaction times are plotted for the different information values of control signals in the MOTOR (red) and the TASK (green) experiments.

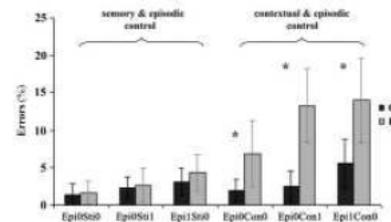


Fig. 4 Percentages of errors (mean \pm SD) in patients (grey) and controls (black). The percentages are plotted for the different information values of control signals. * $P < 0.001$.

increased ($P = 0.0045$), whereas controls performed equally in the different conditions ($P > 0.05$) (Fig. 4). Patients' performance, however, did not deteriorate differently from controls' as the demands of sensory and episodic controls increased in the MOTOR experiment (both interactions Group*Epi: $P > 0.05$) indicating that varying the amount of information conveyed by sensory and episodic signals did not increase patients' error percentages more than in the control group. Finally, comparing EPs in the first and second half of blocks within each group, we found that no effect significantly varied over episodes, (two-tailed t -tests, all $t < 1.9$, $df = 24$, $P > 0.05$) indicating that there was no learning effect within the block for either patients or control subjects.

Regression analyses

Impairments associated with an increased demand of contextual control (i.e. contextual score) significantly predicted

the disorganization score ($R^2 = 0.21$, $P < 0.05$) and in particular, its formal thought disorders subscore ($R^2 = 0.21$, $P < 0.05$) (Fig. 5). The more patients were disorganized or specifically exhibited thought disorders, the higher their EPs when demands of contextual control varied from 0 to 1 bit. On the other hand, SANS and SAPS scores were not found to be predicted by any cognitive performance nor by any combination of cognitive performances. This was also true when the cognitive performance scores were transformed.

Discussion

In a previous study, Koechlin *et al.* (2003) showed that cognitive control is structured on three separate hierarchically organized levels, the corresponding signals of which are treated in turn (sensory > contextual > episodic), thus permitting the selection of the appropriate response given a specific stimulus. In that multistage model, the processing carried out at each level of the control hierarchy is constantly informed by the processing driven by progressively higher levels, giving rise to a *cascade* of top-down, successive influences along the antero-posterior frontal axis. As suggested in introduction, the Koechlin *et al.*'s model is broadly derived from classical theories of executive control (Baddeley and Hitch, 1974) and of the prefrontal cortex organization—in particular those that focus on top-down attentional control, such as the top-down attentional supervisor from Shallice or Passingham's model of 'attention to action' (Passingham, 1993; Shallice, 1998; see also Koechlin and Summerfield, 2007).

Using a paradigm based on this cascade model, we selectively assessed cognitive control in a group of patients with schizophrenia, by evaluating the contributions of each level (sensory, contextual, episodic levels) of the hierarchy to a visuomotor association task. We first showed that the

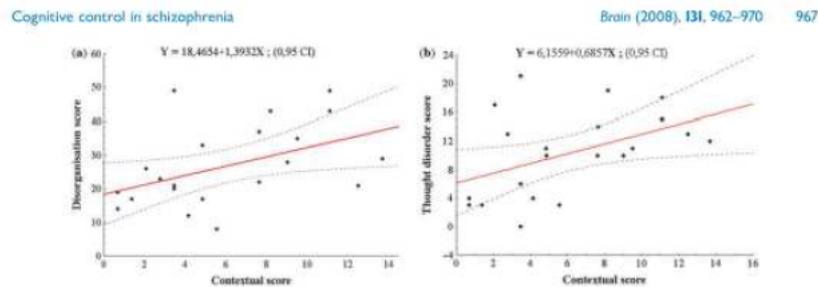


Fig. 5 The linear regression lines (and their respective equations) derived from the linear regressions analyses between patients' contextual score (explanatory factor) and their disorganization score (a) and thought disorder score (b) are shown in red. The 95% confidence interval (CI) around the regression lines are shown in grey.

architecture of cognitive control in schizophrenic patients was roughly similar to that of comparison participants, with a multistage, cascading organization of information processing. Indeed, as in healthy participants, both patients' RTs and EPs were found to increase when processes of control involved progressively higher level stages (e.g. sensory > contextual > episodic control), just as they performed worse as the information conveyed by the control signals increased (e.g. from $I=0$ to 1 bit).

However, despite a similar pattern of increased latencies in both groups, patients' EPs differed from healthy participants with respect to both the level and the demand of cognitive control. Interestingly, this perturbation appeared under certain conditions only. Indeed, patients with schizophrenia were found to perform as well as healthy subjects in tasks requiring sensory and episodic control, whereas they showed great difficulty with tasks requiring the control of contextual cues.

In the MOTOR experiment, both the episodic information—which pre-activated a set of visuomotor associations congruent with the ongoing episode—and the sensory information—which ensured the selection of a particular response among that set—were correctly processed by patients. Furthermore, patients were influenced in the same way as comparison participants by the varying demands of sensory and episodic controls (Epi0Sti0 > Epi0Sti1 and Epi0Sti0 > Epi1Sti0) since neither the main effect of condition nor the interactions between group and conditions were significant. Taken together, these observations confirm previous results obtained by Posada and Franck (2002) who, using a visuomotor association task, showed that patients with schizophrenia correctly used a rule to associate a colour with a response.

On the other hand, patients' performance significantly differed from healthy participants when control of contextual cues was specifically required in order to associate a response with a stimulus (TASK experiment).

This impaired performance could not result from variations in memory load (i.e. from maintaining instructions related to cues over subsequent episodes) because these variations were larger in the two-forced versus single-forced conditions, yet patients performed equally well in both conditions. Error patterns also cannot be explained by patients having forgotten the rules, since these were given to subjects prior to each episode, nor by difficulties in maintaining a task-set in working memory within an episode because EPs did not vary within episodes. Finally, regarding the specificity of the deficit, the poor performance of patients in contextual conditions is unlikely to be attributed to the difficulty of the task. Indeed, healthy subjects were found to perform at comparable levels across the MOTOR (episodic and sensory) and TASK (episodic and contextual) experiments, suggesting an equal discriminating power between both tasks (Chapman and Chapman, 1973b, 2001). This observation is consistent with a specific deficit in context processing where, all else being equal, patients' performances significantly decreased as the task required, specifically, controlling contextual information, whereas those of healthy subjects remained constant.

Context processing impairments have long been considered as a core feature of schizophrenia (Cohen and Servan-Schreiber, 1992) and numerous studies confirm the extent of this deficit in the disease (Stratta *et al.*, 1998; Braver and Cohen, 1999; Stratta *et al.*, 2000; Elvevag *et al.*, 2000; Barch *et al.*, 2003). Patients with schizophrenia tend to preferentially select the most frequent meaning of an ambiguous word and conversely, to neglect information concerning the immediate context, even when this is crucial for selecting a less dominant but more relevant meaning of a word (Cohen and Servan-Schreiber, 1992). Similarly, increased interference effects created by the higher conflict between irrelevant words and relevant contextual information like colour in the Stroop task, is well-documented in patients with schizophrenia (Perlstein *et al.*, 1998, for a review).

As pointed out by Park and collaborators (Park *et al.*, 2003; see also Hemsley and Phil, 2005), context proves to be a highly composite construct, with various dimensions referring to separate processes which may be differentially impaired—some dimensions being possibly intact, while some others may not be. The context multidimensionality may thus render its operationalization subject to some confusion, hence the importance of a detailed and rigorous definition as to what this construct refers to. In a broad sense, context can be defined as an internal representation of any task-relevant information that can be used to mediate an appropriate behavioural response (Braver and Cohen, 1999). Under this account, context may include various things like the prior stimulus, the results of processing a sequence, the task instructions ('episodic context') and can even be extended to physical features of the stimulus itself ('perceptual context': location, size, colour...) (Park *et al.*, 2003). In the light of our results, we propose that not all aspects of this information of context are dysfunctional in schizophrenia. In particular, the episodic dimension of this information could be spared in schizophrenic patients, as suggested by their good performance on the MOTOR experiment where the selection of the appropriate response required correctly and continuously processing episodic context information (task instructions). On the other hand, patients were found to perform worse in conditions in which the selection of the appropriate task-set required taking into account the perceptual, immediate context associated with the target stimulus (the letters' colour). In such tasks, patients' poor performance revealed a default in their perceptual context processing required to select the appropriate task-set among competitive ones (lower/upper case or consonant/vowel discrimination).

As revealed by Koechlin *et al.*'s model (Koechlin *et al.*, 2003), contextual control of action is implemented within the caudal part of the lateral prefrontal cortex (cPF) which mostly overlaps with the well-known Broca's area (BA 44, 45). Recently, the contribution of this region to behavioural control was clarified, when it was shown to be involved in the hierarchical organization of behaviour (Koechlin and Jubault, 2006). Broca's area has been shown to contain a system of executive processes that control the nesting of functional segments that combine in hierarchically organized action plans (Koechlin and Jubault, 2006). In the present study, contextual control impairment observed in patients—e.g. disturbance of a particular level of the control hierarchy—could therefore reveal a more specific problem in the hierarchical organization of action representations that would thus impact on the selection of appropriate task-sets.

Obviously, the selection of inadequate action representations may cause some difficulty in planning and organizing adapted behaviours over time. Consistently, we found that impaired contextual control performance was specifically associated with disorganization symptoms as revealed by

significant regression scores between the contextual score and both the disorganization score and its formal thought disorders subscore. Although significant, it is noteworthy that this association was of a moderate strength ($R^2 = 0.21$), indicating that only a small proportion of variance in disorganization scores was captured by the deficit we observed. As already suggested by Cohen *et al.* in a previous study (Cohen *et al.*, 1999), this could be due to the fact that the disorganization scale we computed includes several subscales, which may themselves be differentially related to context processing deficits.

Taken together, these results are consistent with a number of previous studies using behavioural settings or neuroimaging techniques. The severity of the disorganization syndrome has been many times associated with the extent of executive deficits (Mahurin *et al.*, 1998; Kravartiti *et al.*, 2005). In particular, the profile of performances we observed in the schizophrenia sample is particularly consistent with Cohen *et al.*'s previous reports of a close relationship between disturbances in the processing of context and symptoms of formal thought disorder (Cohen *et al.*, 1999; see also Kerns and Berenbaum, 2003). Our data also fit with several fMRI studies showing that disorganized patients tended to have lower activation in prefrontal regions (BA 9/46, Perlstein *et al.*, 2001; Snitz *et al.*, 2005), the regions that, as predicted by the cascading model, precisely mediate the control of contextual information (Fig. 1).

In the present study, this association further suggests that disorganized behaviour can be partially explained by patients' inability to specifically use contextual information for concomitantly selecting appropriate behavioural representations among competitive ones. This interpretation is strengthened by the fact that this impairment worsened as the influence that contextual cues potentially play in that selection, increased. Finally, given the critical role of Broca's area in the hierarchical organization of human language (Dominey *et al.*, 2003; Musso *et al.*, 2003), impairment at this level of the control hierarchy may also account for clinical manifestations such as disorganized speech. In the light of the cascading model (Koechlin *et al.*, 2003; Koechlin and Jubault, 2006), unexpected topic switches, derailment or tangential responses might indeed arise from an inability to continuously monitor the hierarchical organization of human language, e.g. to coordinate linguistic segments that compose speech in relation to the superordinate goals and subgoals of discourse. As the functional integrity of Broca's area in schizophrenia is still a matter of debate, with previous neuroimaging studies reporting either relatively normal (Perlstein *et al.*, 2001; MacDonald *et al.*, 2005) or reduced Broca's activation (Stevens *et al.*, 1998; Snitz *et al.*, 2005), this hypothesis should be taken cautiously, however. In future work, the use of neuroimaging techniques should allow us to test this assumption more directly.

Conclusion

Patients with schizophrenia do not suffer from general cognitive control impairment but rather from a deficit of some control processes. While episodic and sensory controls were found to be spared, contextual control was specifically impaired. In the present study, this is reflected in an impairment in the selection of an appropriate task-set, given a specific perceptual context. Contextual control can be considered as a set of executive processes involved in the hierarchical organization of behaviour (Koechlin and Jubault, 2006). Patients' deficit in cognitive control thus reflects a specific problem in the hierarchical control of action, leading to the selection of inappropriate behavioural representations for current action plans. Moreover, this impairment was a good predictor of disorganization syndrome score, further suggesting that these clinical manifestations might result, at least partially, from a deficit in the combination or selection of hierarchically organized action representations in the motor and, possibly, the verbal domains. Obviously, such a deficit is likely to impact on patients' social functioning. Indeed, adapted social behaviour and interactions crucially depend on the ability to organize actions in the context of both our own internal, but also external goals inferred from other people's behaviour. We therefore believe that bringing to light the mechanisms underlying such ability may provide, in future, valuable tools to account more efficiently for difficulties in social adaptation that patients with schizophrenia encounter in their everyday life.

Acknowledgements

This research was supported by a grant of the Conseil Scientifique de Recherche, Le Vinatier (CSRA 05). V. Chambon was supported by a scholarship from the French Ministry for Research. The authors report no competing interests. We gratefully acknowledge Mrs Ghislaine Bailly and Prof. Jean-Louis Terra for their help in the inclusion of patients. We wish to thank Coralie Chevallier and Victoria Southgate for their assistance in proofreading this paper as well as the three anonymous reviewers for their helpful comments and suggestions.

References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, DSM-IV, 4th edn., Washington, DC: APA; 1994.
Andreasen NC. The scale for the assessment of negative symptoms (SANS). Iowa City: The University of Iowa; 1983.
Andreasen NC. The scale for the assessment of positive symptoms (SAPS). Iowa City: The University of Iowa; 1984.
Axelrod BN, Goldman RS, Heaton RK, Curtis G, Thompson LL, Chelune GJ, et al. Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *J Clin Exp Neuropsychol* 1996; 18: 338–42.
Baddeley AD, Hitch G. Working memory. In: Bower GA, editor. Recent advances in learning and motivation. Vol. 8. New York: Academic Press; 1974. p. 47–90.

Barch DM, Carter CM, MacDonald AW, Beaver TS, Cohen JD. Context processing deficits in schizophrenia: diagnostic specificity, four week course and relationship to clinical symptoms. *J Abnorm Psychol* 2003; 112: 132–43.
Beaver TS, Cohen JD. Dopamine, cognitive control, and schizophrenia: the gating model. *Prog Brain Res* 1999; 121: 327–49.
Chapman LJ, Chapman JP. Problems in the measurement of cognitive deficit. *Psychol Bull* 1973b; 79: 380–5.
Chapman LJ, Chapman JP. Commentary on two articles concerning generalized and specific cognitive deficits. *J Abnorm Psychol* 2001; 110: 31–9. Review.
Cohen JD, Servan-Schreiber D. Context, cortex and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychol Rev* 1992; 99: 45–77. Review.
Cohen JD, Barch DM, Carter C, Servan-Schreiber D. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. *J Abnorm Psychol* 1999; 108: 120–33.
Dominey P, Hoen M, Blanc JM, Lelekov-Boissard T. Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. *Brain Lang* 2003; 86: 207–25.
Elvevag B, Duncan L, McKenna P, Pravo JP, Rousseaux M. Control functions of the frontal lobes. Modularity of the central-supervisory system? *Cortex* 1999; 35: 1–20.
Gras-Vincendon A, Danion JM, Grange D, Billik M, Willard-Schroeder D, Sichel JP, et al. Explicit memory, repetition priming and cognitive skill learning in schizophrenia. *Schizophr Res* 1994; 13: 117–26.
Hemley DR, Paul M. The Schizophrenic experience: taken out of context? *Schizophr Bull* 2005; 31: 43–53.
Jones SH, Hemley DR, Gray JA. Contextual effects on choice reaction time and accuracy in acute and chronic schizophrenics. Impairment in selective attention or in the influence of prior learning? *Br J Psychiatry* 1991; 159: 415–21.
Kerns JG, Berensbaum H. The relationship between formal thought disorder and executive functioning component processes. *J Abnorm Psychol* 2003; 112: 339–52.
Koechlin E, Jubault T. Broca's area and the hierarchical organization of human behavior. *Neuron* 2006; 50: 963–74.
Koechlin E, Ody C, Kouneifer F. The architecture of cognitive control in the human prefrontal cortex. *Science* 2003; 302: 1181–4.
Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci* 2007; 11: 229–35.
Kravatin E, Dixon T, Frith C, Murray R, McGuire P. Association of symptoms and executive function in schizophrenia and bipolar disorder. *Schizophr Res* 2005; 74: 221–31.
Lezak MD. Newer contributions to the neuropsychological assessment of executive functions. *J Head Trauma Rehabil* 1993; 8: 24–31.
MacDonald AWJrd, Carter CS, Kerns JG, Ursu S, Barch DM, Holmes AJ, et al. Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *Am J Psychiatry* 2005; 162: 473–84.
Mahurin RK, Velligan DM, Miller AL. Executive-frontal lobe cognitive dysfunction in schizophrenia: a symptom subtype analysis. *Psychiatry Res* 1998; 79: 139–49.
Musso M, Moro A, Glauche V, Rijntjes M, Reichenbach J, Buchel C, et al. Broca's area and the language instinct. *Nat Neurosci* 2003; 6: 774–81.
Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971; 9: 97–113.
O'Leary DS, Flaum M, Kesler ML, Flashman LA, Arndt S, Andreasen NC. Cognitive correlates of the negative, disorganized, and psychotic symptoms dimensions of schizophrenia. *J Neuropsychiatry Clin Neurosci* 2000; 12: 4–15.
Park S, Lee J, Folley B, Kim J. Schizophrenia: putting context in context. *Behav Brain Sci* 2003; 26: 96–9.

Passingham RE. The frontal lobes and voluntary action. Oxford: Oxford University Press; 1993.
Perlbstein WM, Carter CS, Barch DM, Baized JW. The Stroop task and attention deficits in schizophrenia: a critical evaluation of card and single-trial Stroop methodologies. *Neuropsychology* 1998; 12: 414–25.
Perlbstein WM, Carter CS, Noll DC, Cohen JD. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am J Psychiatry* 2001; 158: 1105–13.
Posada A, Franck N. Use and automation of a rule in schizophrenia. *Psychiatry Res* 2002; 109: 289–96.
Shallice T, Burgess PW, Frith CD. Can the neuropsychological case-study approach be applied to schizophrenia? *Psychol Med* 1991; 21: 661–73.
Shallice T. From neuropsychology to mental structure. Cambridge: Cambridge University Press; 1998.
Shannon CE. A mathematical theory of communication. *Bell Sys Tech J* 1948; 27: 379–423.

Snitz BE, MacDonald A 3rd, Cohen JD, Cho RY, Becker T, Carter CS. Lateral and medial hypofrontality in first-episode schizophrenia: functional activity in a medication-naïve state and effects of short-term atypical antipsychotic treatment. *Am J Psychiatry* 2005; 162: 2322–9.
Stevens AA, Goldman-Rakic PS, Gore JC, Fulbright RK, Wexler BE. Cortical dysfunction in schizophrenia during auditory word and tone working memory demonstrated by functional magnetic resonance imaging. *Arch Gen Psychiatry* 1998; 55: 1097–103.
Stratta P, Daneluzzo E, Bustini M, Casacchia M, Rossi A. Schizophrenic deficit in the processing of context. *Arch Gen Psychiatry* 1998; 55: 186–8.
Stratta P, Daneluzzo E, Bustini M, Prosperini P, Rossi A. Processing of context information in schizophrenia: relation to clinical symptoms and WCST performance. *Schizophr Res* 2000; 44: 57–67.
Velligan DM, Bow-Thomas CC. [Review]. Executive function in schizophrenia. *Semin Clin Neuropsychiatry* 1999; 4: 24–33.
Wechsler D. Wechsler Adult Intelligence Scale. (3rd edn). Administration and Scoring Manual. San Antonio, TX: Psychological Corporation/Harcourt Brace; 1997.

4. Discussion (1)

L'hypothèse-clé du modèle de Koechlin et coll. (2003) s'articule autour d'une distinction cruciale entre processus de contrôle opérant selon le *contexte perceptif* ou l'*épisode temporel* dans lequel le sujet agit. Ces processus joueraient un rôle essentiel dans la planification, l'organisation et le contrôle *on-line* des comportements volontaires. Nous avons fait l'hypothèse qu'une altération localisée de l'un ou l'autre de ces niveaux de contrôle pouvait rendre compte des troubles du fonctionnement exécutif observés dans la schizophrénie.

Les résultats de la présente étude suggèrent que ces troubles pourraient dépendre d'une atteinte plus fondamentale des processus de traitement et de contrôle des informations de contexte (Chambon et al., 2008). Les patients schizophrènes présentent en effet d'importantes difficultés dans les conditions qui requièrent l'exploitation contrôlée d'indices contextuels – les niveaux de contrôle sensoriel et épisodique étant quant à eux relativement préservés. Ces difficultés se traduisent chez les patients par une incapacité à sélectionner les ensembles d'associations Stimulus-Réponse (droite si Consonne, gauche si Voyelle, par exemple) contextuellement pertinents, c'est-à-dire adaptés aux signaux contextuels entrants (la couleur des lettres).

Le contrôle contextuel est connu pour jouer un rôle crucial dans l'organisation hiérarchique du comportement : de notre capacité à exploiter les informations de contexte environnant dépend notre aptitude à sélectionner, parmi un ensemble d'actions concurrentes, les actions adaptées à ce contexte. Naturellement, la sélection d'actions contextuellement inappropriées peut causer certaines difficultés à planifier ou organiser les comportements dans le temps, c'est-à-dire à ordonner *hiérarchiquement* les différents segments qui constituent la séquence comportementale en passe d'être exécutée (Koechlin & Jubault, 2006). Nous avons trouvé précisément que cette difficulté était, chez les patients, spécifiquement associée aux symptômes de désorganisation (pour des observations similaires, voir Kravariti et al., 2005). Cette observation invite à considérer les manifestations cliniques de la désorganisation comme résultant d'une incapacité, au moins partielle, à coordonner entre elles – à *prioritiser* – les différentes unités motrices qui composent la structure du comportement, moteur (séquence d'exécution de segments moteurs) comme verbal (séquence d'énonciation de segments linguistiques ou articulatoires).

Au plan organique, ce trouble spécifique de l'exploitation des informations contextuelles pourrait traduire l'existence d'anomalies cérébrales localisées, en particulier dans la région caudale du cortex préfrontal latéral (CPL) dont l'activité est en effet étroitement associée au contrôle des informations de contexte (Koechlin et al., 2003). Ces deux dernières décennies, le cortex préfrontal a été, chez le patient schizophrène, un objet d'investigation privilégié. La plupart des études rapporte à ce titre l'existence de dysfonctionnements majeurs dans de nombreuses régions du CPL, en lien avec des perturbations du traitement des informations contextuelles (Barch et al., 2001 ; MacDonald et al., 2003), des difficultés à manipuler l'information stockée en Mémoire de Travail (Cannon et al., 2005 ; Tan et al., 2006) ou à restituer convenablement cette information depuis la mémoire épisodique (Barch et al., 2002 ; Hofer et al., 2003).

En dépit de ces avancées significatives, néanmoins, l'architecture fonctionnelle du contrôle cognitif dans le CPL des patients schizophrènes reste encore mal spécifiée. Nous avons donc adapté, en collaboration avec Guillaume Barbalat, le protocole décrit ci-dessus en neuro-imagerie fonctionnelle (IRMf) chez un groupe de 15 nouveaux patients

schizophrènes et 15 sujets témoins. Nos prédictions étaient les suivantes : pour peu que les résultats de la première étude se confirment, nous nous attendions à observer chez les patients une modulation anormale de la région caudale du CPL, en association avec un trouble pour le contrôle des informations de contexte. Or, cette atteinte pour le niveau de contrôle contextuel pourrait ne pas être aussi localisée qu'elle y paraît, puisqu'elle persiste, voire est aggravée, lorsque que la tâche sollicite un contrôle épisodique accru (Chambon et al., 2008). D'autres régions de la hiérarchie corticale, dont la partie la plus antérieure de cette hiérarchie (région rostrale, BA 10), pourraient donc être impliquées. L'étude que nous avons conduit revêtait par conséquent un second intérêt : celui de déterminer si les troubles observés en matière d'intégration contextuelle témoignaient d'un dysfonctionnement circonscrit à la région dédiée au traitement du contexte (partie caudale), ou s'ils étaient redevables d'un problème de connectivité au sein de la hiérarchie de contrôle elle-même, entre les régions rostrales (épisodique) et caudales (contexte) – comme cela a été par ailleurs récemment suggéré (Yoon et al., 2008).

Organization of Cognitive Control within the Lateral Prefrontal Cortex in Schizophrenia

Guillaume Barbalat, Valerian Chambon, Nicolas Franck, Etienne Koechlin & Chloé Farrer
(Archives of General Psychiatry)

ORIGINAL ARTICLE

Organization of Cognitive Control Within the Lateral Prefrontal Cortex in Schizophrenia

Guillaume Barbalat, MSc, MD; Valerian Chambon, MSc; Nicolas Franck, MD, PhD; Etienne Kochlin, PhD; Chloé Farrer, PhD

Context: Cognitive control is highly affected in schizophrenia, but its overall functional architecture remains poorly understood. A recent study demonstrated that, in healthy subjects, cognitive control is functionally organized within the lateral prefrontal cortex (LPFC) as a cascade of representations ranging from premotor to anterior LPFC regions according to stimuli, the present perceptual context, and the temporal episode in which stimuli occur.

Objective: To determine the functional hierarchical organization of cognitive control within the LPFC in patients with schizophrenia.

Design: Case-control study.

Setting: Hospital-based research units.

Participants: Fifteen schizophrenic patients and 14 controls.

Main Outcome Measures: Behavioral performance and regional brain activity as measured by functional magnetic resonance imaging during a task, varying the amount of information conveyed by episodic and contextual signals.

Results: In patients and healthy controls, activity in caudal LPFC regions varied as episodic and contextual signals, whereas rostral LPFC regions only exhibited an episodic effect. However, patients made more errors than controls when information conveyed by contextual and episodic signals increased. These impairments were related to hypoactivation in caudal LPFC regions and hyperactivation in rostral LPFC regions, respectively. Activation in caudal LPFC regions negatively correlated with the disorganization syndrome score of patients.

Conclusions: In schizophrenic patients, the architecture of cognitive control follows the cascading organization from rostral LPFC regions to caudal LPFC and premotor regions depending on the temporal framing of action and events. We found, however, that immediate contextual signals insufficiently bias the caudal LPFC activity required to select the appropriate behavioral representation. This specific deficit could thus alter the internal consistency of schizophrenic patients' behavior. To compensate for this weakening of contextual influence, schizophrenic patients may inefficiently use temporal episodic information through higher activation in rostral LPFC regions.

Arch Gen Psychiatry. 2009;66(4):377-386

DISTURBANCES OF COGNITIVE control, the ability to coordinate thoughts and actions in relation to internal goals, are robustly associated with schizophrenia and are thought to play a key role in patients' maladaptive and perseverative behaviors.^{1,2} The neural substrates of these deficits have been investigated by numerous studies that have focused on top-down attentional control,^{3,4} working memory,⁵ and episodic memory processes.⁶ In general, these studies reported consistent dysfunctions in various regions of the lateral prefrontal cortex (LPFC) related to impairments in the processing of immediate contextual information,^{7,8} manipulation of information within working memory,^{9,11} and information retrieval from episodic memory.¹²⁻¹⁴ Despite these notable advances, func-

tional divisions of cognitive control within schizophrenic patients' LPFC remain poorly understood. In particular, researchers have failed to clearly describe (1) how the different temporal modules of cognitive control (particularly those related to episode and context) interact and are functionally organized within specialized LPFC subsystems and (2) how dysfunctions in those areas could be related to clinical symptoms.¹⁵

To investigate the overall organization of cognitive control within the LPFC in schizophrenia, we used a unified modular model proposed by Kochlin et al¹⁶ in which representations are distributed in the LPFC according to their temporal structure rather than their content or internal complexity. Specifically, Kochlin et al showed that the LPFC is organized as a cascade of executive processes (from

Author Affiliations: Centre de Neurosciences Cognitives, Université de Lyon, CNRS (Drs Barbalat, Franck, and Farrer, and Mr Chambon), and Centre Hospitalier le Vinatier (Dr Franck), Lyon, France; and Inserm, Université Pierre et Marie Curie and Ecole Normale Supérieure, Paris, France (Dr Kochlin).

(REPRINTED) ARCH GEN PSYCHIATRY/VOL 66 (NO. 4), APR 2009 WWW.ARCHGENPSYCHIATRY.COM

377

Downloaded from www.archgenpsychiatry.com on April 7, 2009
©2009 American Medical Association. All rights reserved.

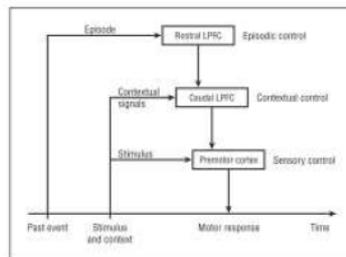


Figure 1. The cascading architecture of cognitive control within the lateral prefrontal cortex (LPFC). The multistage organization of information processing includes a sensory control level involved in selecting the motor responses that are the most appropriate to specific stimuli. This control is subserved by the lateral premotor regions (typically, Brodmann area [BA] 6). The contextual control level is involved in selecting premotor representations (that is, stimulus-response associations) according to contextual signals that accompany the occurrence of stimuli. This control is subserved by the caudal part of the LPFC (typically, BA 9/44/45). The episodic control level is involved in selecting caudal LPFC representations (task sets or consistent sets of stimulus-response associations evoked in the same context) according to the temporal episode in which stimuli occur. This control is subserved by the rostral part of the LPFC (typically, BA 46/10). Reprinted with permission from the American Association for the Advancement of Science.¹⁶

premotor to anterior LPFC regions) that control behavior according to stimuli, the present perceptual context, and the temporal episode in which stimuli occur (**Figure 1**). Additionally, this model takes into account quantitative variations in the demands of these controls, using a mathematical framework that is based on information theory.^{16,17}

The cascading nature of this model is derived from the idea that each stage maintains active representations that are controlled by higher stages, which exert control on representations in lower stages, thus giving rise to a series of top-down, successive processes involved in controlling the appropriate stimulus-response association. This model offers an account of both top-down attentional and fractionation theories of LPFC function, consistent with findings from various domains within cognitive neuroscience, such as attentional control, working memory, and episodic memory retrieval.¹⁶ The cascade model is supported by a wide variety of evidence from neuroimaging studies in humans and lesion studies in nonhuman primates, including those referring to other hierarchical models of branching cognition.^{16,24} We believe that the cascade model provides a useful holistic framework in which to investigate the temporal organization of cognitive control within the LPFC in schizophrenia. Using a protocol inspired by this model, Chambon and coworkers²⁵ previously showed that, in schizophrenic patients, a specific deficit occurred when the selection of an appropriate action was biased by immediate contextual cues. This defective contextual control was shown to account for significant variance in disorganization syndrome scores in patients. These results led to the hypothesis that caudal LPFC, the region of the

prefrontal cortex that subserves contextual control, is dysfunctional in schizophrenia and that this dysfunction could be related to the disorganization syndrome.

To test this hypothesis, we used functional magnetic resonance imaging (fMRI) to evaluate LPFC activation in schizophrenic patients and matched healthy participants during a task that modeled contextual and episodic controls. This task was adapted from the experimental paradigm of Koehlin et al.¹⁶ In the healthy group, we first expected that the increasing demands of contextual and episodic controls would have additive cumulative effects on both behavioral reaction times and local brain activations that gradually add up from rostral to caudal LPFC and premotor regions.¹⁶ Second, as has been shown in previous studies, we expected that the increasing demand of both contextual and episodic controls would enhance error rates in schizophrenic patients.²⁵⁻²⁷ Third, as mentioned, we predicted that there would be insufficient modulation of activity in caudal LPFC regions relative to impaired control of contextual information and an association between caudal LPFC regions' dysfunction and the disorganization syndrome in the patient group. Finally, as patients' LPFC activation was shown to reflect a complex interaction between task difficulty and the subjects' motivation to perform the task,^{28,29} we also sought to determine the nature of any differences in regional brain activation when comparing patients and controls with equivalent behavioral accuracy.

METHODS

PARTICIPANTS

Fifteen schizophrenic patients and 15 healthy controls, who were all right-handed (Edinburgh Handedness Survey³⁰) and matched for age, sex, and years of education, were recruited to participate in the fMRI experiment (**Table 1**). These participants were different from those who had participated in Chambon and colleagues'²⁵ previous study. After the study was completely described to the participants, written informed consent was obtained, as approved by the local ethics committee. All of the participants were paid for their participation. Diagnosis was confirmed for each patient by an MD- and PhD-level clinical psychiatrist (masked to task performance) based on the Structured Clinical Interview of the DSM-IV-TR.³¹ The clinical state of each patient was assessed using the Scale for the Assessment of Positive Symptoms³² and the Scale for the Assessment of Negative Symptoms³³ on the day of testing. Symptoms were grouped into 3 syndromes: reality distortion, psychomotor poverty, and disorganization (Table 1). All patients were clinically stable and treated with only atypical antipsychotic medications. None of the participants had a history of brain trauma, seizure disorder, electroconvulsive therapy, mental retardation, affective disorder, substance abuse, or substance dependence within the past 6 months. In addition to these exclusion criteria, special exclusion criteria for the controls included having a history of an Axis I disorder, having a first-degree relative with a psychotic disorder, and receiving treatment with any psychotropic medication within the past 6 months. One control participant was excluded because of motion artifact (no patients were excluded).

TASK PARADIGM

Subjects had to respond as quickly and accurately as possible to a series of successive colored letters by pressing 1 of 2 re-

Table 1. Clinical and Demographic Characteristics of Schizophrenic Patients and Controls

Characteristic	Mean (SD)		P Value
	Schizophrenic Patients (n=15)	Controls (n=14)	
Male sex, No. (%)	8 (53)	8 (57)	.88
Age, y	35 (10.5)	36 (10.6)	.79
Education, y	11 (1.3)	11 (1.9)	.82
Right-handedness	0.86 (0.09)	0.84 (0.11)	.50
Duration of illness, y	10 (9)		
SANS score	43 (19)		
SAPS score	23 (21)		
Reality distortion score ^a	8 (10)		
Psychomotor poverty score ^b	34 (18)		
Disorganization score ^c	23 (15)		
Chlorpromazine equivalent, ^d mg/d	247 (190)		

Abbreviations: SANS, Scale for the Assessment of Negative Symptoms; SAPS, Scale for the Assessment of Positive Symptoms.
^aSum of the scores for hallucinations and delusions from SAPS.
^bSum of the scores for poverty of speech, flat affect, anhedonia/associality, and avolition from the SANS.
^cSum of the scores for formal thought disorder and bizarre behavior from the SAPS and the score for attention from the SANS.^{11,12}
^dDoprot doses and daily oral atypical antipsychotic drugs at the time of the examination (risperidone in 6 patients, olanzapine in 3 patients, amisulpride in 3 patients, and aripiprazole in 2 patients) were converted to average daily chlorpromazine-equivalent doses using guidelines described by Woods.¹³
 None of the patients received a concurrent typical antipsychotic, anticholinergic agent, sedative treatment, mood stabilizer, antidepressant, or other psychotropic agent.

response buttons (held in the right and left hands). The experimental design was administered using an 8 × 8 Latin square design consisting of 8 series of stimuli (scanning sessions), each presented in 8 separate blocks (behavioral episodes). A Latin square design was used to control for order of presentation of the blocks and transitions between the blocks. Each block included a series of 12 successive stimuli (duration, 500 milliseconds; onset asynchrony, 3500 milliseconds) preceded by an instruction cue (episodic signal) that lasted 4200 milliseconds. Participants learned the instructions before running the experiment to avoid possible biases due to learning effects during the test session. In each scanning session, the 8 blocks formed 4 distinct experimental conditions that crossed the demands of contextual and episodic controls varied by manipulating the contextual (I_{con}) and the episodic (I_{ep}) factors, respectively. The stimulations were delivered with the software Presentation (Neurobehavioral Systems, Albany, California).

Instruction cues varied from block to block and informed the subjects to make quick responses to colored letters or to withhold a response to a no-go stimulus. Contextual control was manipulated by requiring subjects to make either a lower-/upper-case (task 1) or a consonant/vowel (task 2) letter judgment (using the left and right response buttons), the task being determined by the color of the letter (ie, the contextual signal within each block). Episodic control was manipulated by varying the contingencies that linked contextual signals and task sets (Figure 2).

In each block, sequences of contextual signals were pseudorandomized so that 33% of displayed letters were to be ignored. In dual task-set blocks, the ratio of task 1 to task 2 trials was 1. The proportions of 2 successive trials that included identical contextual signals were constant across blocks. Finally, letters were pseudorandomly chosen so that, in each block, the ratio of left- to right-button responses was 1 and the ratio of

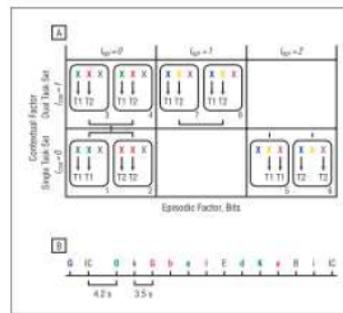


Figure 2. Experimental design of the task paradigm. A, Rounded boxes represent behavioral episodes (1–8) with related stimuli (letters) and instructions. Episodes formed 4 distinct experimental conditions that crossed the episodic factor with the contextual factor. According to the color of the letter (contextual signal), subjects either ignored the letter or performed a vowel/consonant (T1) or lower-/upper-case (T2) discrimination task on the letters. For example, in block 1, contextual signals were either green or white. White signals indicated that subjects should ignore the letter. Green signals indicated that subjects should perform task T1 (single task-set episode). When contextual control was low, the task remained the same across the entire block (T1 or T2, single task-set blocks, $I_{con}=0$ bits, blocks 1, 2, 5, and 8); in high-contextual control blocks, the task changed from trial to trial (T1 and T2, dual task-set blocks, $I_{con}=1$ bit, blocks 3, 4, 7, and 8). Episodic control was manipulated by varying the contingencies linking contextual signals and task sets. When the colors involved in the blocks were green, red, and white, the same colors always denoted the same tasks (in blocks 1, 2, 3, and 4, green always denoted T1, red always denoted T2, and white was always “no-go,” $I_{ep}=0$ bits). However, when the colors were blue, purple, and yellow, the tasks demanded by each color varied on a block-by-block basis (in blocks 5, 6, 7, and 8, blue, purple, and yellow could all denote T1, T2, or no-go, $I_{ep}>0$ bits). B, Typical episode. IC indicates instruction cues; I_{con} and I_{ep} , contextual and episodic factors.

congruent to incongruent letters (same vs different responses for task 1 and task 2) was 1. Accordingly, sensorimotor control was constant across the experiment.

MRI PROCEDURES AND PREPROCESSING

Images were collected using an 1.5-T MRI system. The fMRI blood oxygenation level–dependent signal was measured using a T2*-weighted echo-planar sequence (repetition time, 2500 milliseconds; echo time, 60 milliseconds; flip angle, 90°). Twenty-six axial slices (thickness, 4 mm; gap, 0.4 mm; field of view, 220 mm; matrix size, 64 × 64; in-plane resolution, 3.4 × 3.4 mm²) were acquired per volume. Following functional image acquisition, a high-resolution T1-weighted anatomical image (repetition time, 1970 milliseconds; echo time, 3.93 milliseconds; 256 × 256 matrix; resolution, 1 × 1 × 1 mm³) was collected for each subject.

Image preprocessing was performed using SPM5 (Wellcome Department of Imaging Neuroscience, University College London, London, England). For each subject, each of the 8 scanning sessions contained 155 functional volumes after the first 5 scans were rejected to eliminate the non-equilibrium effects of magnetization. All functional volumes were realigned to the first volume to correct for interscan movement. Functional and structural images were coregistered and transformed¹⁷ into a standardized, stereotaxic space (Montreal Neurological Institute template).¹⁸ Functional data were then smoothed with a 10-mm full-width-at-half-maximum, isotro-

Table 2. Behavioral Performance in Each Group of Subjects and in Matched Performance Subgroups

Condition	Measure	Mean (SD)	
		Controls	Schizophrenic Patients
$I_{con}=0$ bits	Error rate, %	0.9 (1.6)	5.5 (4.3)
	Reaction time, ms	1005 (177)	1137 (310)
$I_{con}=1$ bit	Error rate, %	2.0 (2.1)	9.5 (8.2)
	Reaction time, ms	1369 (190)	1507 (278)
$I_{con}=2$ bits	Error rate, %	3.3 (2.8)	17.5 (13.6)
	Reaction time, ms	1468 (186)	1608 (278)
$I_{epi}=0$ bits	Error rate, %	7.3 (5.7)	20.2 (12.5)
	Reaction time, ms	1352 (194)	1538 (283)

Abbreviations: I_{con} and I_{epi} , contextual and episodic factors.

pic gaussian kernel and temporally processed in a high-pass filter with a frequency cutoff of 128 seconds. Serial correlations were accounted for by use of an autoregressive model of the first order. To control for possible noise artifacts in the data, we used a weighted least-squares approach, in which we down-weighted images with high noise variance.³⁹

STATISTICAL ANALYSIS

Error rates and reaction times for correct trials acquired during scanning were analyzed using analyses of covariance with subject as a random factor, group as a between-subject factor, context as a within-subject factor, and episode as a within-subject covariate. When significant, interactions were decomposed using *t* tests. To conduct these analyses, we used the statistical software STATISTICA7 (StatSoft Inc, Tulsa, Oklahoma).

For the fMRI data, we first conducted voxelwise exploratory analyses of frontal regions subserving each level of cognitive control (context and episode) in both the schizophrenia and control groups. Using SPM5, statistical parametric *t* score maps were computed from local fMRI signals using a linear multiple regression analysis with conditions (modeled as box-car functions convolved by the canonical hemodynamic response function) and scanning series as covariates.⁴⁰ For all conditions, we defined the preparation phase as the interval between the instruction cue and the presentation of the first stimulus, and the execution phase as the period from the first stimulus until the end of the series of stimuli. In the current study, we analyzed the frontal regions that were engaged in cognitive control exertion during the execution phase. Specifically, regions that demonstrate a contextual effect are those that have higher activation in the dual ($I_{con}=1$ bit) than in the single ($I_{con}=0$ bits) task-set episodes, with $I_{epi}=0$ bits. Regions that demonstrate an episodic effect are regions whose activation parametrically vary with I_{epi} . In a second level of analysis, contrasts were performed using a random-effect model. In each group, we identified regions that showed a contextual effect and regions that showed an effect of episode, but no contextual effect (computed by masking each region related to the episodic effect with the contextual effect, using an uncorrected voxelwise threshold $P < .05$). Montreal Neurological Institute coordinates were transformed to the standard space of Talairach and Tournoux⁴¹ and reported as *t* scores. In accordance with our a priori hypothesis and on the basis of the known distributed functional and structural anatomy of cognitive control,^{39,42} we restricted our analyses to the lateral frontal cortex by masking with WFU PickAtlas software.⁴³ A voxelwise significance threshold of $P < .05$ was chosen (corrected for the false discovery rate).

We then conducted hypothesis-driven functional regions of interest analyses in the different regions (rostral LPFC, caudal LPFC, and premotor regions) identified by the exploratory analyses in healthy subjects. Activations in the voxel that was the more significant in each of these regions (ie, the peak voxel) were separately entered into univariate repeated-measure analyses of covariance, with subject as a random factor, hemisphere (left vs right) and number of alternatives (single vs dual task set) as within-subject factors, episode ($I_{epi}=0, 1, \text{ or } 2$ bits) as a within-subject covariate, and group (patients vs controls) as a between-subject factor (when performing between-group analyses). When significant, interactions were further assessed using *t* tests. For schizophrenic patients, we also tested correlations between brain activity and the 3 symptom syndromes (reality distortion, psychomotor poverty, and disorganization) using the Pearson test. To conduct these analyses, we used STATISTICA7.

RESULTS

BEHAVIORAL ANALYSIS

Reaction Times

The analysis of covariance performed on reaction times showed significant effects of episode ($F=100.16, P < .001$) and context ($F=197.89, P < .001$), revealing slower reaction times as the demands of cognitive controls increased. Patients' reaction times, however, did not deteriorate in a manner that was distinct from the controls' as the demands of contextual and episodic controls increased (all interactions with group factor: $F < 0.67, P > .05$), indicating that varying the amount of information conveyed by contextual and episodic signals did not increase patients' reaction times more than it did in the control group (Table 2 and Figure 3A and B).

Error Percentages

Participants' error percentages were found to significantly increase with the contextual ($F=4.16, P < .05$) and episodic ($F=66.23, P < .001$) factors. Significant interactions between group and cognitive factors were observed in both the contextual ($F=4.58, P < .05$) and the episodic ($F=13.26, P < .001$) factors. These effects were due to a greater decrement in performance among patients than controls regarding both the episodic and contextual factors (Table 2 and Figure 3C and D). Patients performed worse than controls for $I_{con}=0$ and 1 bit and for $I_{epi}=0, 1, \text{ and } 2$ bits (all $t > 3.50, P < .002$). Finally, by comparing reaction times and error percentages in the first and second parts of episodes, we found that no effect significantly varied across the episodes (all interactions, $F < 1.92, P > .05$), indicating that cognitive control was recurrently exerted across episodes and that there was not a learning effect within the blocks for patients relative to controls.

fMRI ANALYSIS

Exploratory Voxelwise Contrasts

In controls, frontal regions that showed an effect of context were found bilaterally in the caudal LPFC. (Bro-

mann area 9/44/45, inferior/middle frontal gyrus) and premotor cortex (Brodmann area 6, middle frontal gyrus). Frontal regions that exhibited an episodic, but not contextual, effect were found bilaterally in the rostral LPFC (Brodmann area 10/46, inferior/middle frontal gyrus) (Figure 4A and B and Table 3).

In patients, frontal regions that showed a contextual effect were found bilaterally in the caudal LPFC and in the left premotor cortex (however, these peaks were not significant, $P = .09$, false discovery rate-corrected). Frontal regions that exhibited an episodic, but not contextual, effect were found bilaterally in the rostral, caudal, and premotor regions (Figure 4A and B and Table 3).

Functional Region of Interest Analyses

To examine the group effects from the series of exploratory voxelwise contrasts, fMRI signal changes for each subject were extracted for each of the 3 LPFC regions identified in the healthy group (Table 3). In controls as in patients, activations in rostral LPFC regions linearly varied with the episodic factor only ($F > 5.4$, $P < .05$), though in patients, this effect was only observed in the right hemisphere. In both groups, activations in these regions were independent of context ($F < 1.5$, $P > .05$). Moreover, between-group analyses revealed a main group effect ($F = 7.1$, $P = .01$), with patients activating these regions to a greater extent than controls (Figure 4D and E). There was no interaction between group and cognitive factors ($F < 1.2$, $P > .05$).

In controls, activations in the caudal LPFC regions linearly varied with episode ($F = 10.6$, $P < .005$) and context ($F = 19.3$, $P < .001$). In patients, activations in these regions only varied with the episodic factor ($F = 5.0$, $P < .05$), independent of context ($F = 0.8$, $P > .05$). There was no interaction between episode and context in any group ($F < 0.2$, $P > .05$). Moreover, between-group analyses showed a significant group \times context interaction ($F = 9.8$, $P < .005$) (Figure 4C, F, and G) but not a significant group \times episode interaction or a main group effect ($F < 2.9$, $P > .05$). Planned contrasts using t tests revealed that controls activated these regions to a greater extent in the dual task-set condition ($I_{ms} = 1$ bit) than in the single task-set condition ($I_{ms} = 0$ bits) ($t = 3.2$, $P < .005$), whereas patients did not ($t = 0.5$, $P > .05$).

One could argue that the hypoactivation in the patients' caudal LPFC could result from a bias of the analysis we used because we localized this region of interest from activations found in the healthy group alone. We therefore conducted a between-group analysis on the caudal LPFC activations whose localizations were different in each group. That is, peak-voxel activations for each subject in the healthy group were extracted from the caudal LPFC regions that were specifically identified by the contextual contrast in this group. Conversely, peak-voxel activations for each subject in the schizophrenic group were extracted from the caudal LPFC regions that were specifically identified by the contextual contrast in this latter group ($P < .09$, false discovery rate-corrected). Then, these activations were entered into a new analysis of covariance, which still revealed a group \times contextual interaction ($F = 13.5$, $P < .001$) but no main effect of group or an interaction between group and episode ($F < 1.15$, $P > .05$).

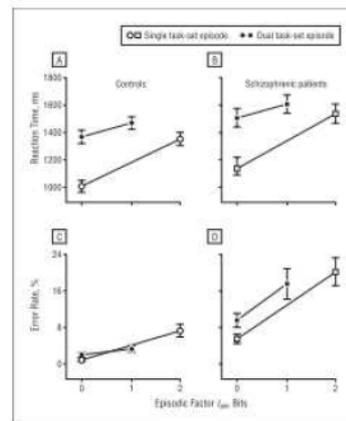


Figure 3. Behavioral results of the task paradigm. Reaction times (A and B) and error rates (C and D) (mean \pm standard error across participants) across experimental conditions.

Region of interest-based analysis performed on activations in premotor regions showed significant effects of episode ($F = 4.0$, $P < .05$) and context ($F = 11.5$, $P = .001$) in controls. In schizophrenic patients, we observed a significant episodic ($F = 5.3$, $P < .05$) but not a contextual ($F = 1.7$, $P > .05$) effect. There was no interaction between episodic and contextual factors in any group ($F < 0.1$, $P > .05$). Between-group analyses revealed neither a significant group effect nor significant interactions between group and cognitive factors ($F < 3.0$, $P > .05$) (Figure 4H and I).

Correlation analyses conducted between the 3 syndromes scores (reality distortion, disorganization, and psychomotor poverty) and fMRI signal change in these regions revealed a significant correlation in caudal LPFC regions for disorganization only ($r = -0.59$, $P < .05$). An increased disorganization score was associated with a decreased signal intensity change. Results from other correlation analyses between each of the 3 syndromes and signal intensity change in rostral LPFC and premotor regions were not significant.

It is noteworthy that schizophrenic patients as well as controls were found to complete some of the blocks by chance. If satisfactory blocks were defined as those completed with an accuracy greater than 65%,¹⁶ a mean of 0.75 blocks per run (standard deviation [SD], 1.06) were considered to be completed with chance in the patients, compared with a mean of 0.12 (SD, 0.32) in the control participants ($P < .05$). Such a difference in performance has been argued to constitute a so-called performance bias for the interpretation of the neuroimaging results. Indeed, in that case, differences in activations could be interpreted as resulting from patients' poor en-

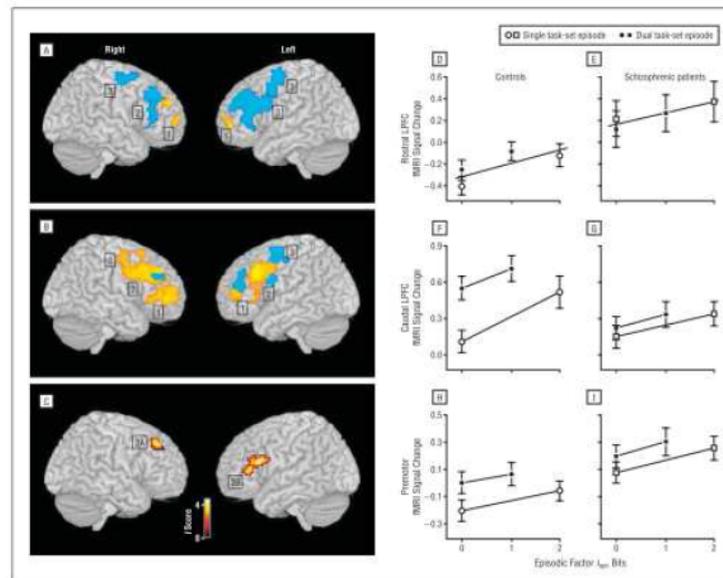


Figure 4. Differences in prefrontal cortex activations between patients with schizophrenia and healthy controls. Topography of brain activation in healthy controls (A) and in patients with schizophrenia (B) (blue indicates regions exhibiting a contextual effect, yellow, regions showing an episodic but no contextual effect; 1, rostral lateral prefrontal cortex [LPFC]; 2, caudal LPFC; 3, premotor cortex). C, Brain regions demonstrating contextual effects that were significantly greater in healthy control participants than in participants with schizophrenia (2A indicates dorsocaudal LPFC; 2B, ventrocaudal LPFC; D-1, Peak-voxel activations (averaged regression coefficients \pm standard errors across participants) in LPFC regions involved in cognitive control in healthy subjects are plotted against the episodic factor. Left and right activations are averaged together. fMRI indicates functional magnetic resonance imaging.

engagement in the task, rather than resulting from a specific cognitive deficit.

One way to control this bias is to remove blocks for which performance is unsatisfactory. When we reran the analysis considering only blocks for which the accuracy was acceptable (ie, accuracy >0.65), we found that there were no behavioral differences between the 2 groups regarding either episode or context ($F < 0.21$, $P > .05$). However, we still found a group effect in rostral LPFC regions ($F = 6.97$, $P < .05$), with patients activating this region more than controls. Likewise, caudal LPFC regions still demonstrated a group \times context interaction ($F = 3.76$, $P = .05$), with patients showing no modulation of activation related to the contextual factor in these regions. Activation in the caudal LPFC was only found to correlate negatively with the disorganization score (Pearson $r = -0.54$, $P < .05$). Finally, there were no differences in fMRI signals in premotor regions between patients and controls.

Another way to address the potential confounding of task performance and group is to match subgroups of pa-

tients and controls based on performance.¹⁵ We therefore performed additional analyses with subgroups of subjects matched for behavioral accuracy (8 patients and 7 controls, $t = 1.72$, $P > .05$) (Figure 5A-D). These analyses generally replicated those already described. In particular, even after selecting subjects with similar performance patterns, patients still underactivated the caudal LPFC regions compared with controls regarding the contextual factor (group \times context interaction: $F = 5.2$, $P < .05$) (Figure 5G and H). Additionally, patients matched for performance had higher activation levels than controls in rostral LPFC regions (main group effect: $F = 5.9$, $P < .05$) (Figure 5E and F). No other effect involving group was significant in any of the 3 regions of interest (Figure 5).

COMMENT

We examined the hierarchical organization of cognitive control within the LPFC based on the temporal framing

Table 3. Within-Group Activation in Frontal Regions Displaying Episodic and Contextual Effects

Lateral Frontal Cortex Region	Estimated BA	Coordinates ^a			t Score ^b	Volume, mm ³	FDR-Corrected P Value
		x	y	z			
Controls							
Contextual effect ^c							
Left middle frontal gyrus, caudal PFC	9	-42	39	36	6.27	37 084 ^d	.04
Left precentral gyrus, premotor cortex	6	-39	0	45	5.74	37 084 ^d	.04
Left inferior frontal gyrus, caudal PFC	44	-45	15	21	5.40	37 084 ^d	.04
Right middle frontal gyrus, caudal PFC	9	42	33	39	5.21	8277 ^d	.04
Right inferior frontal gyrus, caudal PFC	45	30	27	3	4.78	8277 ^d	.04
Right superior frontal gyrus, premotor cortex	6	27	-9	54	5.16	3422	.04
Episodic effect ^e							
Left superior frontal gyrus, rostral PFC	10	-27	54	-3	4.23	185	.04
Right middle frontal gyrus, rostral PFC	46	39	48	30	4.06	740	.04
Right middle frontal gyrus, rostral PFC	10	33	63	9	3.49	139	.04
Schizophrenic Patients							
Contextual effect ^c							
Left middle frontal gyrus, caudal PFC	9	-33	42	12	4.92	2867	.09
Left precentral gyrus, premotor cortex	6	-57	-3	24	4.67	2081	.09
Right middle frontal gyrus, caudal PFC	9	33	36	27	4.41	786	.09
Left precentral gyrus, premotor cortex	6	-45	0	54	4.39	2127	.09
Left superior frontal gyrus, premotor cortex	6	-21	-6	57	4.35	1988	.09
Episodic effect ^e							
Right middle frontal gyrus, caudal PFC	9	42	42	33	5.52	23 814 ^d	.03
Right superior frontal gyrus, premotor cortex	6	30	-9	45	3.36	23 814 ^d	.03
Left middle frontal gyrus, caudal PFC	9	-42	21	33	5.09	14 843	.03
Right middle frontal gyrus, rostral cortex	10	27	51	0	4.77	8046	.03
Left middle frontal gyrus, rostral cortex	46	-36	48	9	4.02	1295	.03
Left superior frontal gyrus, premotor cortex	6	-33	-9	45	3.39	647	.03

Abbreviations: BA, Brodmann area; FDR, false discovery rate; PFC, prefrontal cortex.
^aFrom the stereotaxic atlas of Talairach and Tournoux.⁴¹
^bRegional peak activation representing blood oxygenation-level dependent signal change that reached a threshold of $P < .05$ (corrected for the FDR) in a random-effect analysis.
^cThese peaks belong to the same cluster of activation.
^dThese peaks belong to the same cluster of activation.
^eExcluding contextual effect.
^fThese peaks are nonsignificant but are reported because we wanted to show that the activations are not absent in schizophrenic patients regarding the contextual effect.
^gThese peaks belong to the same cluster of activation.

of action and events in a sample of patients with schizophrenia and a group of matched healthy subjects. We first showed that, in healthy controls, the architecture of cognitive control is organized as a multistage, cascading organization of information processing along a rostrocaudal axis of the LPFC, which is consistent with previous results.^{16,25} In the comparison group, activity in the rostral LPFC only resulted from an effect of episode, whereas activity in the caudal LPFC resulted from effects of both context and episode.

In this cascading architecture, however, schizophrenic patients demonstrated dysfunctional recruitment of specialized areas involved in controlling episodic and contextual information to guide the selection of the appropriate action. Indeed, schizophrenic patients made more errors than controls when information conveyed by both contextual and episodic signals increased (consistent with our previous study²⁵), which was associated with lower activation in the caudal LPFC and higher activation in the rostral LPFC, respectively. Finally, activation in the caudal LPFC—the region subserving contextual control—was associated, as expected, with the disorganization syndrome in the patients' group.

Two methodological issues have typical implications for interpreting specific cognitive functions thought to be impaired in schizophrenia. First, medications used to treat schizophrenia may influence behavioral performance and brain function. However, our findings of both reduced activation in caudal LPFC regions and enhanced activation in rostral LPFC regions in patients are consistent with findings of other studies that included schizophrenic patients who were not taking neuroleptic drugs.^{2,21,42} Therefore, the current pattern of results is likely not due to the treatment patients received.

A second common methodological problem in interpretation of fMRI data in schizophrenic patients is that poor performance may confound changes in functional brain activation.^{26,29} The specific patterns of patient-control differences in prefrontal cortex activation found in this study persisted, however, when compared with conditions in which patients' behavioral performances were matched with those of controls. Thus, the pattern of functional brain activation we observed is unlikely to result from a failure to engage in the task and rather expresses some inherent disturbances in schizophrenia.

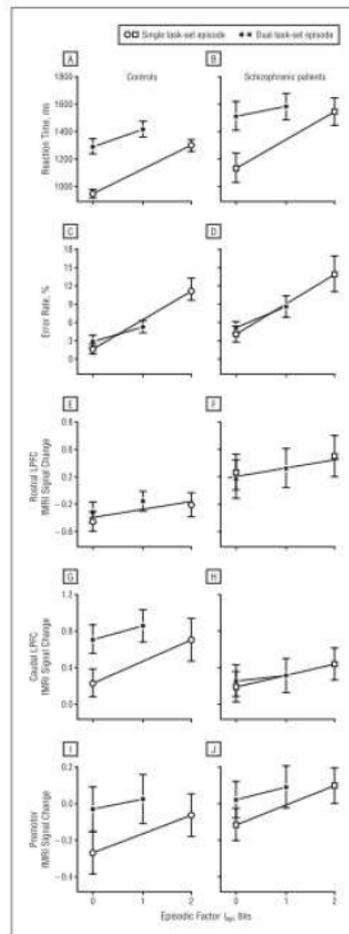


Figure 5. Behavioral and neuroimaging differences between patients with schizophrenia and healthy controls matched for performance. A-D, Behavioral results (mean \pm standard error across participants). E-J, Factorial analyses of regional activations. Peak voxel activations (averaged regression coefficients \pm standard errors across participants) in lateral prefrontal cortex (LPFC) regions involved in cognitive control in healthy subjects are plotted against the episodic factor. fMRI indicates functional magnetic resonance imaging.

It is worth noting that the box-car design of our study may at least partly confound additional cognitive processes that may have contributed to the differential effects observed between patients and controls.³⁶ In particular, the single and dual task-set conditions do not only differ in the demands on contextual control but also in the working memory load during the corresponding task blocks. Indeed, dual task-set blocks require subjects to maintain twice as many stimulus-response mappings (4 vs 2) in mind for the response selection. However, if patients' impaired performances were the consequence of impaired maintenance of contextual signals within working memory, the increasing contextual demand ($L_{in} = 0$ bits to $L_{in} = 1$ bit) should have had a differential effect on reaction times or error percentages in patients compared with controls between the first and second parts of behavioral episodes, which was not the case.

Results from our study support numerous previous findings from top-down attentional theories of cognitive control,⁷ indicating an association between context-processing impairment, the disorganization syndrome, and dysfunction of caudal LPFC (though only its dorsal part was often found to be disturbed) in schizophrenia.^{7,8,31,45,47-49} Whereas researchers have provided convincing accounts for dysfunction in this area (eg, inappropriate dopaminergic modulation from subcortical systems),^{27,50} they have not clearly described the critical role this region plays in some of the major cognitive and clinical disturbances in schizophrenia.¹⁵ We believe that our study provides such clarification. Indeed, the cascade model considers that the caudal LPFC is functionally organized as a set of processes involved in control of the hierarchical structures of action plans (eg, lower-/upper-case or consonant/vowel discrimination tasks) according to immediate contextual signals (eg, the color of the letter).^{18,51} Therefore, caudal LPFC dysfunction likely affects selection of the appropriate behavioral representation, which might complicate the planning and organization of contextually adapted behaviors. Consequently, this specific problem in the hierarchical monitoring of action could contribute to disharmonious behaviors (ie, maladaptive or perseverative behaviors), which is further corroborated by the negative correlation that we found between the disorganization syndrome score and patients' caudal LPFC activity.

A number of previous investigators have proposed other models for a better characterization of the physiopathologic substrates of impaired cognitive control in schizophrenia. One of those suggests that disturbances in the ability to detect conflict or errors in ongoing processing, which may be due to the function of the anterior cingulate cortex, may lead to deficits in selective attention, ie, the ability to enhance the processing of task-relevant information.^{43,52-55} Such abnormal selective attentional functioning in schizophrenia has been demonstrated by numerous studies using the Stroop task, in which the participant is required to name the color of a stimulus word while ignoring its meaning (eg, the word red printed in blue ink).⁵⁶ Interestingly, in the Stroop task, the color of the word could be assimilated to an immediate contextual signal as defined by the cascade model. Indeed, as an intrinsic dimension of the stimulus, the color

is thought to be involved in the control of lower-order, more automatic responses (ie, the meaning of the word). Impaired anterior cingulate conflict monitoring could therefore represent a potential cause for the weakening of contextual influences in schizophrenia,⁴⁹ a hypothesis that should be further investigated in the future.

Another well-known paradigm that shows impaired cognitive control in schizophrenia uses the continuous performance task, in which subjects are presented with a sequence of letters and are instructed to respond to a prespecified probe (X) only if it follows a particular contextual cue (A).^{7,8,31,47,57} In this AX continuous performance task, context processing broadly refers to the ability to represent and maintain task-relevant information in working memory necessary to appropriately process the subsequent target. In this paradigm, contextual control mostly requires overcoming response interference on BX trials, which occur after a very high proportion of A cues. In both the AX continuous performance task and the cascade model, contextual control requires managing conflict between competing stimulus-response mappings but does not involve the maintenance and management of multiple task sets in memory, as is the case in the episodic condition of the present study. In the cascade model, context more closely refers to information associated with the immediate, physical features of the target itself (eg, letter's color), whereas more temporal processes (eg, task instructions) are devoted to episodic control.¹⁹ Based on the cascade model, our previous study revealed that impaired context processing in schizophrenia is specifically related to immediate task-relevant information, the more temporally distant information being adequately controlled in this illness.²⁵

Investigations of LPFC functioning in schizophrenia related to the working memory model have often revealed impaired dorsal LPFC activity (eg, in relation to impaired context processing) but normal^{11,58} or enhanced^{5,10} ventral LPFC activity. In contrast, we showed an impaired recruitment of contextual control by the caudal LPFC in both dorsal and ventral sectors (Figure 4C). Although we cannot formally exclude the possibility that the rather long duration of illness in our patient sample did not contribute to decreases in their ventral LPFC activation,^{5,10} these differences may be justified by variations in the study tasks. Indeed, the cascade model makes no claims about the functional ventrodorsal segregation of the LPFC,¹⁹ whereas tasks that use the working memory paradigm can usually reveal such functional dissociation based on the modality or content of processed information (eg, representation of contextual cues in dorsal LPFC vs phonological storage in the ventral LPFC).^{39,60} Findings from studies that used the working memory model could thus bring into focus the functional organization of the posterior LPFC in schizophrenia. Within this region, which is globally ineffective at controlling immediate contextual signals, patients may have impaired specialization in the dorsal LPFC related to higher-order information processing, and they may try to compensate for this deficit through ventral LPFC activation.⁹

A possible alternative compensatory mechanism is suggested by the higher activation in rostral LPFC—the region subserving episodic control—observed in patients

with schizophrenia despite their impaired behavioral performances while controlling episodic signals. Higher activation in schizophrenic patients' LPFC in the context of normal or impaired performances has often been categorized as cortical inefficiency (graphically represented by an inverted U-shaped function shifted to the left), especially in rostral LPFC regions.^{20,61,62} However, the present results, together with others that have shown reduced dorsal LPFC but enhanced rostral and ventral LPFC activations, suggest an alternative account: cognitive compensation, rather than a mere shift on the inverted U curve.^{9,10,31} Indeed, while the cascade model claims that these regions are involved in selecting caudal LPFC representations, such hyperactivation could be interpreted as a consequence of the additional, though inefficient, effort that patients may expend to retrieve the poorly integrated contextual information.^{21,63,64} This interpretation is corroborated by a previous study from MacDonald et al³¹ that demonstrates that, instead of using context processing, schizophrenic patients may use an inefficient encoding and retrieval episodic strategy—related to enhanced activation in rostral LPFC. In this line of arguments, the disruption of episodic control observed in schizophrenic patients could be the result of an inappropriate binding process between temporal and more immediate aspects of information^{65,66} that may be due to a primary inefficient encoding strategy of contextual cues.

Submitted for Publication: March 28, 2008; final revision received October 30, 2008; accepted November 2, 2008.

Correspondence: Guillaume Barbalat, MSc, MD, Centre de Neurosciences Cognitives, UMR 5229 CNRS, 67 Pinel Blvd, 69675 Bron Cedex, France (gbarbalat@isc.cnrs.fr).

Financial Disclosure: None reported.

Funding/Support: This research was supported by a grant from the Conseil Scientifique de la Recherche, Le Vinatier. Mr Barbalat was supported by grant DEA20050904971 from the Fondation pour la Recherche Médicale.

Additional Contributions: We wish to thank Philippe Domenech, MSc, and Chrystèle Ody, PhD, for their assistance on the data analysis as well as the 4 anonymous reviewers for their helpful comments and suggestions.

REFERENCES

1. Mahurin RK, Veligan DI, Miller AL. Executive-frontal lobe cognitive dysfunction in schizophrenia: a symptom subtype analysis. *Psychiatry Res*. 1998;79(2):139-149.
2. Kirovski E, Dixon T, Frith C, Murray R, McGuire P. Association of symptoms and executive function in schizophrenia and bipolar disorder. *Schizophr Res*. 2005;74(2-3):221-231.
3. Shallice T. *From Neuropsychology to Mental Structure*. Cambridge, England: Cambridge University Press; 1988.
4. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*. 2001;24:167-202.
5. Baddeley A. The episodic buffer: a new component of working memory? *Trends Cogn Sci*. 2000;4(11):417-420.
6. Tulving E. Episodic memory: from mind to brain. *Annu Rev Psychol*. 2002;53:1-25.
7. MacDonald AW III, Carter CS. Event-related fMRI study of context processing in dorsolateral prefrontal cortex of patients with schizophrenia. *J Abnorm Psychol*. 2003;112(4):669-687.
8. Barch DM, Carter CS, Brown TS, Sabb FW, MacDonald A III, Noll DC, Cohen JD. Selective deficits in prefrontal cortex function in medication-naïve patients with schizophrenia. *Arch Gen Psychiatry*. 2001;58(3):280-288.
9. Tan H, Sust S, Buchholz JW, Mattay VS, Meyer-Lindenberg A, Egan MF, Wein-

berger DR, Callicott JH. Dysfunctional prefrontal regional specialization and compensation in schizophrenia. *Am J Psychiatry*. 2006;163(11):1969-1977.

10. Tan HY, Choo W, Fong CSL, Chia RWL. fMRI study of maintenance and manipulation processes within working memory in first-episode schizophrenia. *Am J Psychiatry*. 2005;162(10):1848-1858.

11. Carion TD, Glahn DC, Kim J, Van Epp TDM, Karlsgodt K, Cohen MS, Nuechterlein KH, Bava S, Shrivastava D. Dorsolateral prefrontal cortex activity during maintenance and manipulation of information in working memory in patients with schizophrenia. *Arch Gen Psychiatry*. 2005;62(10):1071-1080.

12. Hoffer A, Weiss EM, Golaszewski SM, Siedentopf CM, Brinkhoff C, Kormsner C, Falber S, Fritschbacher WW. Neural correlates of episodic encoding and recognition of words in unmedicated patients during an acute episode of schizophrenia: a functional MRI study. *Am J Psychiatry*. 2003;160(10):1802-1808.

13. Hoffer A, Weiss EM, Golaszewski SM, Siedentopf CM, Brinkhoff C, Kormsner C, Falber S, Fritschbacher WW. Neural correlates of episodic encoding and recognition of words in unmedicated patients during an acute episode of schizophrenia: a functional MRI study. *Am J Psychiatry*. 2003;160(10):1802-1808.

14. Barch DM, Cosmides JG, Compton T, Snyder AZ. Working and long-term memory deficits in schizophrenia: is there a common prefrontal mechanism? *J Abnorm Psychol*. 2002;111(3):478-494.

15. Barch DM. The cognitive neuroscience of schizophrenia. *Annu Rev Clin Psychol*. 2005;1:321-353.

16. Koechlin E, Ostry D, Koenigler F. The architecture of cognitive control in the human prefrontal cortex. *Science*. 2003;302(5648):1181-1185.

17. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27:379-423, 623-656.

18. Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci*. 2007;11(6):229-235.

19. Sakai K, Passingham RE. Prefrontal interactions reflect future task operations. *Nat Neurosci*. 2003;6(1):75-81.

20. Brewer TS, Reynolds JR, Dussanben D. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*. 2003;39(4):713-726.

21. Assad WF, Rainer G, Miller EK. Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol*. 2000;84(1):451-459.

22. Petrides M, Alvarado B, Evans AC, Meyer E. Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. *Proc Natl Acad Sci U S A*. 1993;90(3):873-877.

23. Baudre D, D'Esposito M. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J Cogn Neurosci*. 2007;19(12):2082-2098.

24. Hazy TE, Frank MJ, O'Reilly RC. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos Trans R Soc Lond B Biol Sci*. 2007;362(1465):1601-1613.

25. Chandon V, Franck N, Koechlin E, Falck E, Ciaparra G, Azorin J, Farrer C. The architecture of cognitive control in schizophrenia. *BRAIN*. 2008;131(pt 4):962-970.

26. Danion JM, Rizzo L, Brunet A. Functional mechanisms underlying impaired recognition memory and conscious awareness in patients with schizophrenia. *Arch Gen Psychiatry*. 1999;56(7):639-644.

27. Cohen JD, Servan-Schreiber D. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychol Rev*. 1992;99(1):45-77.

28. Ramsey NF, Koning HAM, Welles P, Cahn W, van der Linden JA, Kahn RS. Excessive recruitment of neural systems subserving logical reasoning in schizophrenia. *Brain*. 2002;125(pt 6):1793-1807.

29. Callicott JH, Bertolino A, Mattay VS, Langheim FJ, Duyn J, Coppola R, Goldberg TE, Weinberger DR. Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. *Cereb Cortex*. 2000;10(11):1078-1092.

30. Oldfield RC. The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia*. 1971;9(1):97-113.

31. MacDonald AW III, Carter CS, Kerns JG, Unger L, Barch DM, Holmes AJ, Stenger VA, Cohen JD. Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *Am J Psychiatry*. 2005;162(3):479-484.

32. Barch DM, Cosmides JG. Abnormal parietal cortex activation during working memory in schizophrenia: verbal/phonological coding disturbances versus domain-general executive dysfunction. *Am J Psychiatry*. 2007;164(7):1096-1098.

33. Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. *J Clin Psychiatry*. 2003;64(6):663-667.

34. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.

35. Andreasen NC. *The Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City: University of Iowa Press; 1984.

36. Andreasen NC. *The Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City: University of Iowa Press; 1983.

37. Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. *Hum Brain Mapp*. 1995;3(3):165-189.

38. Evans AC, Hamber M, Collins DL, MacDonald D. A MRI-based probabilistic atlas of neuroanatomy. In: Shorvon S, Fish D, Andermann, Bydder GM, Stefan H, eds. *Magnetic Resonance Scanning and Epilepsy*. New York, NY: Springer; 1994:263-274. NATO Science series A.

39. Dierkerichsen J, Stadmeier R. Detecting and adjusting for artifacts in fMRI time series data. *Neuroimage*. 2005;27(3):624-634.

40. Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp*. 2004;2(4):189-210.

41. Talairach JTP. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme Medical Publishers; 1988.

42. Goldman-Rakic PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Plum F, Mountcastle V, eds. *Handbook of Physiology: The Nervous System V*. Bethesda, MD: American Physiological Society; 1987:373-417.

43. Fuster JM. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. New York, NY: Raven Press; 1989.

44. Malfair JA, Laurent PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*. 2003;19(3):1233-1239.

45. Smitz BE, MacDonald A III, Cohen JD, Cho RY, Becker T, Carter CS. Lateral and medial hypofrontality in first-episode schizophrenia: functional activity in a medication-naïve state and effects of short-term atypical antipsychotic treatment. *Am J Psychiatry*. 2005;162(12):2322-2329.

46. Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ. The trouble with cognitive subtraction. *Neuroimage*. 1996;4(2):97-104.

47. Barch DM, Carter CS, MacDonald AW III, Braver TS, Cohen JD. Context-processing deficits in schizophrenia: diagnostic specificity, 4-week course, and relationships to clinical symptoms. *J Abnorm Psychol*. 2002;111(2):132-143.

48. Stratta P, Daneluzzo E, Busini M, Proserpio P, Pisci A. Processing of context information in schizophrenia: relation to clinical symptoms and verbal performance. *Schizophr Res*. 2005;74(1-2):57-67.

49. Perlstein WM, Dux MJ, Carter CS, Noll DC, Cohen JD. Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biol Psychiatry*. 2002;53(1):25-38.

50. Brewer TS, Barch DM, Cohen JD. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol Psychiatry*. 1999;46(3):312-328.

51. Koechlin E, Jubault T. Broca's area and the hierarchical organization of human behavior. *Neuron*. 2006;50(5):963-974.

52. Kerns JG, Cohen JD, MacDonald AW III, Johnson MK, Stenger VA, Alenstein H, Carter CS. Decreased conflict- and error-related activity in the anterior cingulate cortex in subjects with schizophrenia. *Am J Psychiatry*. 2005;162(10):1833-1839.

53. Carter CS, Mintun M, Nichols T, Cohen JD. Anterior cingulate gyrus dysfunction and selective attention deficits in schizophrenia: a [18F]FDG PET study during single-trial Stroop task performance. *Am J Psychiatry*. 1997;154(12):1670-1675.

54. Carter CS, MacDonald AW III, Ross LL, Stenger VA. Anterior cingulate cortex activity and impaired self-monitoring of performance in patients with schizophrenia: an event-related fMRI study. *Am J Psychiatry*. 2001;158(9):1423-1428.

55. Heckers S, Weiss AP, Deckersbach T, Goff DC, Mowbratt RJ, Bush G. Anterior cingulate cortex activation during cognitive interference in schizophrenia. *Am J Psychiatry*. 2004;161(14):2077-2085.

56. Hensk A, Sato R. Schizophrenia and the Stroop effect. *Behav Cogn Neurosci Rev*. 2004;3(1):42-59.

57. Servan-Schreiber D, Cohen JD, Stenger VA. Schizophrenic deficits in the processing of context: a test of a theoretical model. *Arch Gen Psychiatry*. 1996;53(12):1105-1112.

58. Perlstein WM, Carter CS, Noll DC, Cohen JD. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am J Psychiatry*. 2001;158(7):1105-1113.

59. Petrides M. Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philos Trans R Soc Lond B Biol Sci*. 1996;351(1348):1455-1462.

60. Goldman-Rakic PS. The prefrontal landscape: implications of functional architecture for understanding human mental and the central executive. *Philos Trans R Soc Lond B Biol Sci*. 1996;351(1346):1445-1453.

61. Callicott JH, Mattay VS, Verchinski BA, Mattay VS, Egan MF, Weinberger DR. Complexity of prefrontal cortical dysfunction in schizophrenia: more than up or down. *Am J Psychiatry*. 2003;160(12):2209-2218.

62. Glahn DC, Ragland JD, Abramoff A, Barrett J, Loria AR, Baudon CE, Volinger DI. Beyond hypofrontality: a functional meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Hum Brain Mapp*. 2005;25(1):1-9.

63. Welles AP, Schacter DL, Goff DC, Rauch SL, Alpert NM, Fischman AJ, Heckers S. Impaired hippocampal recruitment during normal modulation of memory performance in schizophrenia. *Biol Psychiatry*. 2002;53(1):48-56.

64. Heckers S, Rauch SL, Goff D, Savage CR, Schacter DL, Fischman AJ, Alpert NM. Impaired recruitment of the hippocampus during conscious recollection in schizophrenia. *Nat Neurosci*. 1998;1(4):318-323.

65. Rizzo L, Danion JM, van der Linden M, Grangé O. Patients with schizophrenia remember that an event has occurred, but not when. *Br J Psychiatry*. 1996;168(4):427-431.

66. Burgen F, Marczewski P, Mitchell KJ, van der Linden M, Johnson MK, Danion J, Salamé P. Impaired performance in a working memory binding task in patients with schizophrenia. *Psychiatry Res*. 2004;125(3):247-255.

5. Discussion (2)

5.1 Hypo-activation préfrontale et dysconnectivité rostro-caudale

Cette dernière étude confirme dans leur ensemble les prédictions que nous avons formulées précédemment. Les patients schizophrènes présentent un déficit d'intégration des signaux contextuels qui se traduit, au plan cérébral, par une hypoactivation des régions préfrontales caudales. Chez les patients, cette hypoactivation témoigne de la pauvreté des influences contextuelles sur la sélection de leurs réponses. Or, nous observons que les patients compensent la pauvreté de ces influences *top-down* en se référant de manière excessive aux informations épisodiques de la tâche. C'est ce que semble traduire, chez les patients, l'hyper-activation observée dans la partie rostrale de la hiérarchie de contrôle.

Cette observation invite naturellement à formuler l'hypothèse d'un dysfonctionnement interne à la hiérarchie frontale, dont pourrait rendre compte un problème de connectivité entre les différents modules de cette hiérarchie (Yoon et *al.*, 2008). Pour le déterminer plus précisément, nous avons conduit avec G. Barbalat des analyses de connectivité effective (SEM-PPI) entre les régions caudales et rostrales de la hiérarchie chez les sujets contrôle et les patients schizophrènes de notre échantillon. Les résultats préliminaires de cette analyse confirment nos attentes. La sélection des informations épisodiques est associée, chez les sujets sains, à une augmentation des liens fonctionnels entre les régions rostrales (antérieures) et caudales (postérieures) de la hiérarchie préfrontale. Ces liens, en revanche, sont absents chez les patients schizophrènes (figures 3, p.53, & figure 4, p.54).

Cette absence d'influence des régions antérieures sur les régions plus postérieures de la hiérarchie de contrôle pourrait traduire un problème plus fondamental de *binding* temporel. Chez les patients, en effet, les régions rostrales du CPL n'exercent pas leur fonction habituelle de sélection des représentations caudales. Cette sélection, anormale ou absente chez les patients, pourrait affecter l'intégration des indices de contexte immédiat au sein de l'épisode temporel en cours. L'hypothèse d'un déficit en *binding* temporel a déjà fait l'objet de plusieurs investigations dans la schizophrénie (voir Rizzo et *al.*, 1996 ; Burglen et *al.*, 2004), mais les corrélats cérébraux de ce déficit sont encore mal connus. Les résultats préliminaires de notre analyse de connectivité suggèrent que ce déficit pourrait tabler sur un défaut de connectivité interne aux régions caudales et rostrales du cortex préfrontal dorsolatéral. Ce déficit illustrerait les difficultés qu'ont certains patients schizophrènes – en particulier les patients désorganisés – à intégrer des éléments temporellement distants au sein d'une réponse unique, adaptée au contexte immédiat dans lequel cette réponse doit être produite (Barbalat, Chambon, Ody et *al.*, soumis).

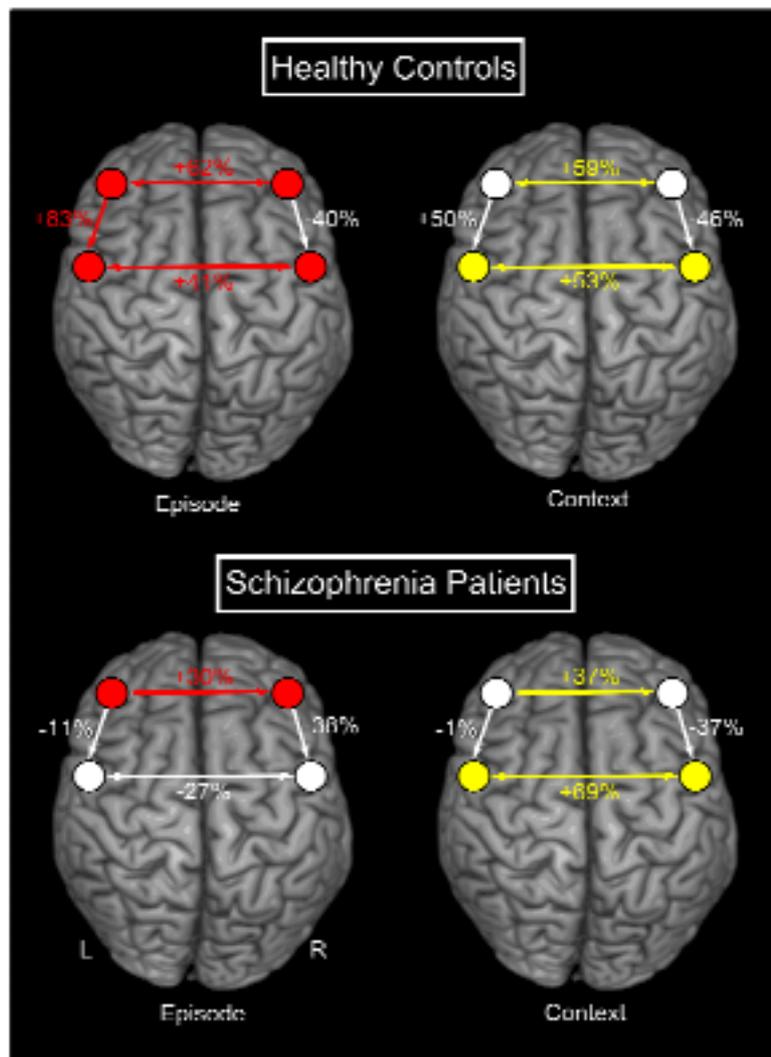


Figure 3. Diagramme des path coefficients entre les régions préfrontales latérales, chez les sujets sains et les patients schizophrènes. Les activations frontales et les path coefficients qui augmentaient significativement avec les facteurs CONTEXTE et EPISODE sont respectivement présentés en jaune et en rouge (NB : les analyses statistiques ont été réalisées sur les variations absolues de ces coefficients, tandis que la figure ne présente que leurs variations relatives).

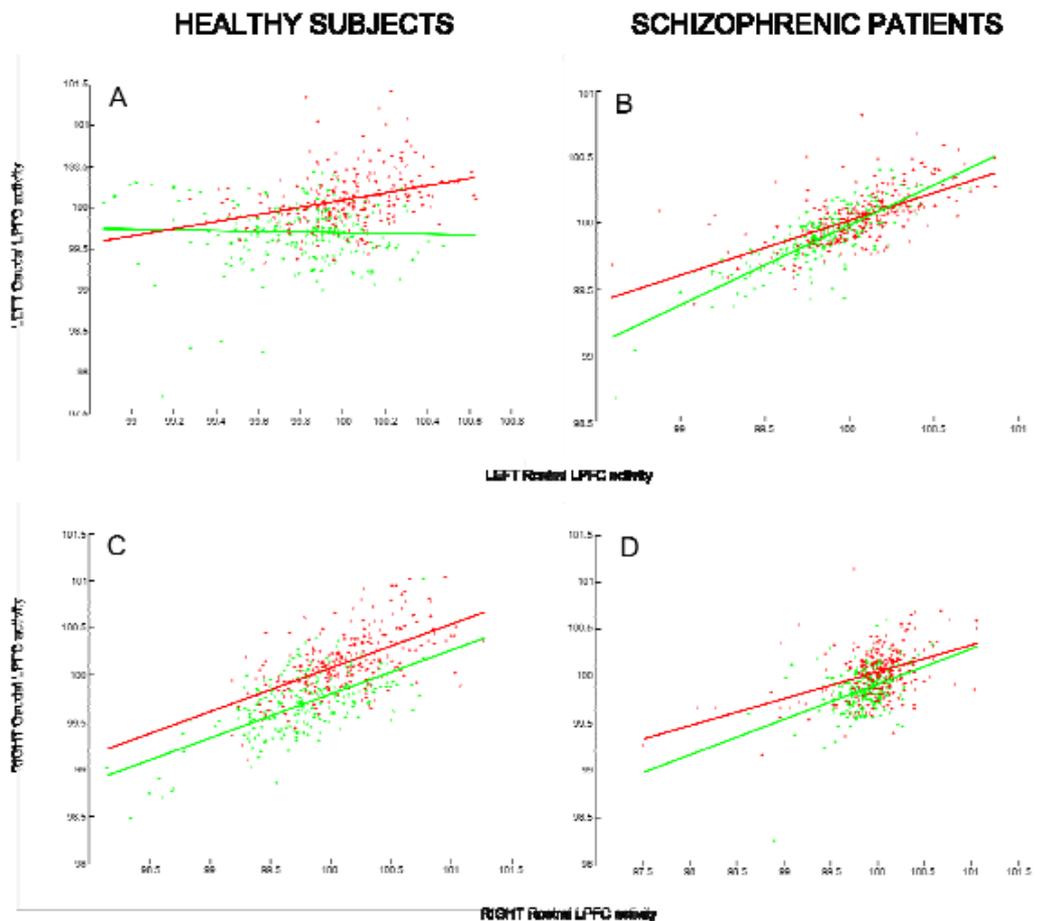


Figure 4. Analyses PPI dans le CPL rostral gauche et droit, chez les sujets sains et les patients schizophrènes. Points verts : valeurs du signal BOLD pour une sollicitation faible de contrôle épisodique (baseline). Points rouges : valeurs du signal BOLD pour une sollicitation élevée de contrôle épisodique (epi). (A & B) L'activité (mean-corrected activity) dans le CPL caudal gauche est présentée comme une fonction de l'activité (mean-corrected activity) dans le CPL rostral gauche (en unités arbitraires). (C & D) L'activité dans le CPL caudal droit est présentée comme une fonction de l'activité dans le CPL rostral droit.

5.2 Du contrôle de l'action à sa compréhension

Comme nous le soulignons en conclusion de notre étude comportementale (Chambon et al., 2008), un déficit en matière d'intégration contextuelle peut avoir une influence dramatique sur le fonctionnement social des patients schizophrènes.

La régulation de nos interactions sociales dépend en effet étroitement de notre capacité à organiser nos actions dans le contexte de nos propres plans et objectifs, mais également

dans le contexte des plans et des objectifs de nos congénères. Une incapacité (ou du moins une capacité réduite) à intégrer, dans nos propres plans d'actions, une représentation adaptée des intentions d'autrui peut favoriser l'élaboration de comportements bizarres, ou dissonants, comme ceux que l'on observe chez certains patients schizophrènes désorganisés. Cette incapacité pourrait également contribuer à l'émergence des troubles de la mentalisation fréquemment rapportés dans cette dimension de la maladie (Zalla et al., 2006).

Plusieurs travaux ont été conduits sur les liens probables que troubles exécutifs et défaut de mentalisation entretiennent dans la schizophrénie : les principaux symptômes productifs, par exemple, pourraient être redevables d'un défaut de contrôle de l'action volontaire, corrélatif d'une altération plus fondamentale de l'intention d'agir (Frith, 1992, 2004, 2005). Si cette hypothèse a donné naissance à une ligne de recherche extrêmement féconde, les résultats demeurent cependant assez controversés (McCabe, 2004, pour revue). Nous verrons dans la seconde partie de ce travail que l'association d'un trouble du contrôle cognitif de l'action avec un déficit en *monitoring* des intentions d'agir (les siennes, ou celles d'autrui) est une hypothèse qui se heurte à des insuffisances d'ordre méthodologique comme théorique. Une manière de résoudre ces insuffisances pourrait consister à explorer ces troubles de la mentalisation à la lumière d'un modèle hiérarchique de la cognition, semblable à celui que nous avons convoqué dans la première partie de ce travail. S'ils ont été longtemps circonscrits à l'étude du seul comportement normal, les modèles hiérarchiques font en effet, depuis peu, l'objet d'un regain d'intérêt pour l'étude des processus psychopathologiques (Fletcher & Frith, 2009, pour revue). Dans ce cadre, nous proposerons de reconsidérer les processus d'attribution intentionnelle comme de simples décisions élaborées « en cascade », et distribuées le long d'une hiérarchie corticale où se croisent les influences de traitements sub- et super-ordonnés. Nous insisterons ensuite sur l'importance de décomposer – comme nous l'avons fait pour le construit théorique des Fonctions Exécutives – les troubles de la mentalisation rapportés dans la schizophrénie en régions *spécifiques* de (probables) dysfonctionnements. Cette parcellisation du trouble implique de redéfinir la notion que les processus de mentalisation ciblent habituellement : l'« intention », la mienne ou celle d'autrui. Plutôt que d'être unique, l'intention est en effet une notion composite susceptible d'être déclinée en plusieurs sous-types, de complexité variable (Pacherie, 2000, 2008). Susceptible, par conséquent, de mobiliser des processus neurocognitifs distincts, pouvant faire l'objet de perturbations spécifiques (Walter et al., 2009). L'étude de ces processus, normaux ou dysfonctionnels, fera précisément l'objet de la seconde partie de ce travail, dont nous proposerons d'interpréter les résultats à la lumière d'un modèle hiérarchique des représentations d'actions intentionnelles en 3^{ème} personne.

Partie 2 Les mécanismes d'attribution d'intentions chez le sujet sain et le patient schizophrène Un modèle hiérarchique des représentations d'actions intentionnelles en 3ème personne

Introduction

La notion de « Théorie de l'Esprit » (ToM) s'est imposée comme une hypothèse de recherche majeure dans le champ des sciences cognitives. Au sens large, cette notion désigne la capacité à expliquer, anticiper ou prédire le comportement d'autrui sur la base des états mentaux qu'on lui prête – croyances, désirs, intentions, etc. Ces processus d'attribution intentionnels apparaissent aujourd'hui nécessaires, puisque lorsqu'ils sont défaillants (comme cela pourrait être le cas dans la schizophrénie), la compréhension de la vie mentale d'autrui s'en trouve perturbée. Or, les mécanismes neurocognitifs qui sous-tendent ces processus d'attribution mentale sont encore aujourd'hui mal connus, quand ils ne font pas l'objet de vives controverses (Van Overwalle & Baetens, 2009, pour revue). La raison en est que la plupart des études concernées négligent le caractère foncièrement composite des états mentaux sur lesquels portent ces processus d'attribution. La notion d'*intention*, en particulier, supporte des niveaux de complexité variable qui, au plan fonctionnel, ne sont pas strictement équivalents (Pacherie, 2000, 2008). Comprendre ou prédire une action observée peut en effet requérir la représentation d'intentions tantôt *motrices* (lecture du but immédiat), tantôt *privées* (représentation du but général de l'action), tantôt *sociales* (sous-jacentes aux actions dirigées vers autrui). En négligeant la diversité de ces *niveaux* intentionnels, les paradigmes expérimentaux actuels laissent souvent indéterminé le détail des mécanismes qui leur sont associés.

Or, la mise en évidence de tels mécanismes pourrait inviter à une relecture féconde de la cognition sociale humaine, et, dans le champ de la psychopathologie, pourrait contribuer à une meilleure compréhension des anomalies de la représentation de soi, de l'autre, et de la distinction entre soi et le monde, caractéristiques de certaines psychoses au premier rang desquelles figure la schizophrénie. Un modèle influent de la schizophrénie suggère en effet qu'une partie des symptômes caractéristiques de cette maladie pourrait être expliquée par l'incapacité des patients à comprendre et se représenter les intentions d'autrui (Frith, 1992). Ce déficit en ToM refléterait, en aval, une incapacité à pouvoir expliquer et prédire les comportements des autres sur la base de leurs intentions supposées.

Nous procéderons tout d'abord à un état des lieux des modèles théoriques élaborés pour rendre compte de cette capacité robuste de « lecture intentionnelle »⁴, puis nous évaluerons l'intérêt et la pertinence de ces modèles pour la compréhension de la clinique schizophrénique. Ces évaluations successives nous conduiront à examiner les insuffisances théoriques de la notion d' « intention », centrale pour la compréhension des troubles explorés. Une redéfinition précise de cette notion conduira à la présentation d'une série d'études comportementales que nous avons menées chez le sujet sain, puis chez le patient schizophrène. La méthode générale de ces études se trouve directement inspirée du formalisme Bayésien, qui, nous le verrons, a déjà fait ses preuves dans de nombreux domaines de la cognition, dont la perception, l'apprentissage, et certaines fonctions cognitives de plus haut niveau (voir Baker et *al.*, 2006, 2007 ; Kilner et *al.*, 2007a, 2007b). Enfin, nous présenterons les résultats préliminaires d'une étude en neuro-imagerie fonctionnelle (IRMf) que nous avons conduit chez le sujet sain et qui adapte les protocoles comportementaux que nous présentons en première partie.

1 Mindreading : comprendre et prédire le comportement.

1.1 De la théorie-théorie aux processus de simulation motrice.

Le champ d'investigation des « cognitions sociales » a pour objet d'étude les mécanismes cognitifs en jeu dans la compréhension mutuelle de soi et d'autrui. Au sens plus restreint du terme, la cognition sociale humaine embrasse tous les processus cognitifs pertinents pour la perception et la compréhension de nos semblables (Blakemore, 2004, pour revue). La capacité proprement humaine de pouvoir comprendre et prédire les comportements de nos congénères pourrait témoigner d'une aptitude plus fondamentale à se représenter ou « lire » l'esprit d'autrui (*mindreading*): c'est en vertu d'états mentaux inobservables (croyances, désirs, intentions, émotions, etc.) qu'Untel a les comportements qu'il manifeste, et c'est parce que je lui attribue de tels états mentaux qu'il m'est dès lors possible d'expliquer ou de prédire les comportements qu'il aura. La compréhension d'autrui dépendrait étroitement de cette aptitude robuste d'attribution et de représentation mentales, automatique, voire irréprouvable, et sans doute d'apparition précoce (Baron-Cohen et *al.*, 1985 ; Leslie, 1987 ; Frith & Frith, 2003).

La nature des mécanismes élémentaires sur lesquels reposerait cette capacité de « lecture mentale », cruciale pour la compréhension d'autrui et de soi-même, a fait et fait encore l'objet de débats très vifs. Selon les tenants de la « Theory-theory » (ou « théorie de la théorie »), comprendre nos semblables nécessiterait une théorie préalable, tacite, du fonctionnement de l'esprit, mobilisant des axiomes et des mécanismes d'inférence élémentaires (Gopnik, 1993 ; Gopnik & Meltzoff, 1994) (v. encadré 1, p.61). Mais les données les plus récentes accumulées dans le champ des neurosciences – et plus particulièrement dans celui de la cognition motrice – ont progressivement opposé aux partisans de la « theory-theory » de solides contre-arguments. La perception des états mentaux d'autrui pourrait dépendre uniquement de la capacité, somme tout naturelle, à

⁴ Ici et dans la suite du texte, nous emploierons l'adjectif « intentionnel » en un sens plus restreint que celui promu initialement par Brentano (1911) : est « intentionnel » un processus, ou une faculté qui *porte sur* ou qui est *à propos* d'une « intention ».

simuler son point de vue (« théorie de la simulation »)(Goldmann, 1995 ; Gordon, 1996 ; Gallese & Goldmann, 2003). L'acquisition humaine des concepts mentaux, comme l'aptitude à comprendre et prédire les conduites intentionnelles de nos semblables, relèverait donc, non pas d'une capacité de théorisation générale, mais d'une capacité plus concrète à répliquer en soi la vie mentale de l'autre – simulation *a minima* qui, en retour, induirait chez le simulateur l'état mental associé au comportement observé. Disposant de représentations motrices similaires à celles que possède l'agent dont je simule l'action, j'accèderais de fait aux intentions sous-jacentes à cette action, et, partant, à l'univers mental de l'agent (Wolpert et al., 2003 ; Blakemore & Decety, 2001 ; Metzinger & Gallese, 2003) (v. encadré 2, p.64).

1. La Théorie de la théorie (« Theory-theory ») – les origines Les tenants de la « Theory-theory » (ou « théorie de la théorie ») postulent que le sujet détermine, se représente et raisonne à propos des états mentaux d'autrui en utilisant une théorie tacite, « naïve », de la structure et du fonctionnement de l'esprit, en partie fondée sur ses expériences passées et sur les lois qu'il en a tirées (Leslie, 87 ; Gopnik, 93). Nos capacités de « mentalisation » engageraient donc des processus inférentiels, intellectuels, plutôt qu'émotionnels ou motivationnels. L'émergence de cette capacité à expliquer (intentionnellement) les comportements en termes de croyances et de désirs dépendrait étroitement des interactions que l'enfant entretient avec son entourage ; ces interactions le conduiraient, progressivement, à faire des hypothèses sur des variables cachées (les croyances ou les désirs des membres de son entourage, par exemple) expliquant les régularités comportementales qu'il observe. A ce stade, l'enfant serait l'équivalent d'un « petit scientifique » (Gopnik & Meltzoff, 1994) testant des hypothèses successives (« il croit que », « il veut que ») le conduisant de proche en proche à élaborer une théorie naïve du fonctionnement psychologique de ses semblables. L'enfant disposerait donc déjà d'un concept de croyance ou de désir, quoique primitif (ses croyances sont toujours vraies et les autres ont toujours les mêmes désirs que lui), qui gagnerait en exactitude au fil des années. La capacité d'attribuer à autrui des intentions ou des croyances propres serait donc liée à la capacité plus élémentaire de pouvoir inférer des états mentaux caractéristiques du comportement observé, et de la situation dans laquelle il se manifeste. La mise en œuvre de ces règles d'inférence serait cruciale pour comprendre autrui, mais également pour se comprendre soi-même (Carruthers, 1996). La connaissance de soi ne serait donc pas immédiate, mais inférentielle, c'est-à-dire médiatisée par des inférences automatiquement réalisées en présence d'êtres biologiques. Il existerait néanmoins une différence de degrés entre la connaissance de soi et celle d'autrui – nous disposerions en effet de plus d'informations sur nous-mêmes dans la mesure où nous serions en meilleure position pour les recueillir –, mais cette différence de degrés ne serait jamais une différence de genre. Dans un cas (se connaître soi-même) comme dans l'autre (connaître autrui), nous appliquerions en effet la même théorie tacite du fonctionnement de l'esprit (Gopnik & Meltzoff, 1994).

L'existence d'un tel système de représentations partagées a été explicitement suggérée par la découverte, chez le singe, de neurones d'un genre particulier, les « neurones miroirs ». Identifiés dans le cortex ventral pré-moteur du singe macaque, cette nouvelle classe de neurones visuo-moteurs est activée de manière identique lorsque l'animal se prépare à exécuter un acte et lorsqu'il voit ce même acte exécuté par l'un de ses congénères (Rizzolatti et al., 1996 ; Gallese et al., 1996 ; Rizzolatti et al., 2000). La décharge des neurones-

miroirs dans le cerveau de l'observateur *simulerait* à proprement parler les mouvements de l'agent perçu ; bien que l'animal ne reproduise pas ouvertement l'action observée, une partie de son système moteur s'active en effet « comme si » il exécutait cette action qu'il observe. En associant automatiquement les mouvements perçus de l'agent avec les représentations motrices contenues dans le lexique moteur de l'observateur, les neurones-miroirs contribueraient ainsi, selon les auteurs, à la formation de « représentations motrices partagées », au sens de représentations communes à celui qui *exécute* (ou a l'intention d'exécuter) et à celui qui *perçoit* l'action (ou qui cherche à en déchiffrer le contenu intentionnel) (Jeannerod, 2001). L'activation de ces représentations motrices ouvrirait l'accès aux buts et aux intentions de l'agent simulé, et, partant, à la possibilité de former des prédictions quant à ses comportements.

Chez l'homme, plusieurs expériences en imagerie fonctionnelle suggèrent également l'existence d'un réseau cortical commun d'activation entre l'observation, l'imagination et l'exécution de l'action (Iacoboni et al., 1999 ; Rizzolatti et al., 2001 ; Grèzes et al., 2003). Ce « système miroir », comprenant plusieurs aires cérébrales, dont le lobule pariétal inférieur, l'aire prémotrice ventrale et le gyrus frontal inférieur, pourrait sous-tendre les processus d'imitation et de lecture intentionnelle⁵. De fait, certains auteurs ont postulé que les neurones polyvalents de ce système-miroir, recrutés de manière identique pour l'observation, la simulation ou la réalisation d'un même acte, supporteraient la capacité de faire lien entre des représentations en *première* et en *troisième* personne, et, partant, inviteraient à une relecture féconde de la cognition sociale humaine (Gallese, 2003).

En effet, qu'est-ce qui caractérise la spécificité des relations que nous entretenons avec nos semblables, sinon la possibilité que nous avons de partager avec eux des actions ou des émotions similaires ? L'élément central de la cognition sociale en général, et de la compréhension d'autrui en particulier, pourrait dépendre de cette capacité du cerveau à faire lien entre des expériences vécues en première et en troisième personne (*i.e.* lien « je fais et je sens » avec « il fait et il sent ») (Wolpert et al., 2003). Le système-miroir humain, recruté à la fois quand nous exécutons une action dirigée vers un but et lorsque nous observons autrui accomplir une action similaire, pourrait supporter cette faculté de liaison naturelle, et sous-tendre ainsi la possibilité d'un rapport *empathique* à l'autre (Georgieff, 2000 ; Singer et al., 2004).

Confortant cette hypothèse, un système-miroir similaire, impliquant cette fois les centres viscéro-moteurs, a été identifié pour l'émotion : des structures identiques (l'insula antérieure et le cortex cingulaire antérieur) semblent contribuer de manière univoque à la perception de stimuli induisant ou exprimant le dégoût (comme des visages) et à l'expérience même du dégoût (Phillips et al., 1997 ; Jabbi et al., 2008). Ces données sont cohérentes avec des études plus anciennes conduites dans le domaine de la perception émotionnelle faciale (Lipps, 1903). Il est en effet admis que la perception d'une expression faciale déclenche chez l'observateur un processus inconscient d'imitation *a minima* de l'expression observée. Cette imitation induit l'état affectif correspondant à l'expression ainsi imitée, et, l'induisant, permet de faire l'expérience directe de l'état mental ressenti par l'autre (Lee et al., 2006).

Ces mécanismes innés d'activation partagée de représentations motrices ou émotionnelles, offrent des arguments convaincants en faveur de l'hypothèse

⁵ La participation de ces régions au système miroir humain fait néanmoins débat et de récentes études suggèrent de circonscrire l'« activité miroir » à la partie rostrale du lobule pariétal (*i.e.* le sulcus intrapariétal antérieur ; Gallese et al., 2004 ; Tunik et al., 2007) tandis que l'activité miroir dans le gyrus frontal inférieur devrait être limitée à sa partie postérieure et pourrait s'étendre jusqu'à « la partie adjacente du cortex prémoteur » (Gallese et al., 2004).

simulationniste. Nos capacités naturelles de simulation - incarnées par l'existence de ces « systèmes-miroirs » de représentations partagées – nous garantiraient en effet un accès privilégié à l'univers mental d'autrui, aux émotions qu'il éprouve et aux intentions motrices qui guident son action, indépendamment de toutes connaissances préalables sur la psychologie humaine ou sur le fonctionnement de l'esprit en général (Blakemore & Decety, 2001 ; Gallese et al., 2003 ; Keysers & Gazzola, 2007).

2. La théorie simulationniste – les origines La théorie simulationniste n'est pas homogène et se décline en deux courants principaux : **2a)** Le courant introspectionniste : selon Goldman (1995), la méthode simulationniste ne se suffit pas à elle-même. Simuler un comportement nécessite en effet une connaissance préalable de ce que « croire » ou « savoir » (par exemple) signifient. Cette connaissance est, toujours selon Goldman, acquise par introspection, et se trouve être la condition de possibilité de processus même de simulation – la raison pour laquelle lorsque je simule autrui, je peux éprouver ce qu'il éprouve lui-même sous la modalité de la « croyance » ou du « savoir ». Dans le cadre de l'introspectionnisme, le sujet demeure la référence de la simulation. Simuler, c'est s'imaginer être dans une situation, sans pour autant y être soi-même. Ce n'est que dans un second temps que le produit de la simulation (une émotion, une motivation, une intention particulières) est projeté sur l'individu dont on souhaite comprendre l'état. **2b)** Le courant anti-introspectionniste : selon Gordon (1996), la simulation est nécessairement neutre quant à la personne qui simule. Simuler, c'est se projeter, non pas dans un contexte fictif, reproduit en soi, mais dans le contexte de l'autre à proprement parler (lequel interagira avec mon propre système de représentations, émotionnelle et motivationnelle). La dimension introspective est évacuée de la simulation, nécessairement radicale : simuler consiste avant tout à simuler une action, un comportement, sans faire appel à quelque concept que ce soit, puis à éprouver, en retour, les valeurs associées à cette action ou à ce comportement. L'enfant, à ce titre, ne simulerait jamais les membres de son entourage comme si ceux-là étaient doués de propriétés psychologiques particulières (du type « croyance », « désir », ou « intention »). Lorsqu'il est confronté à une situation particulière, qui met en scène l'action ou le comportement d'un individu particulier, l'enfant construit la valeur émotionnelle de l'action engagée en la simulant, et conclut que la situation a telle et telle propriété, émotionnelle, motivationnelle, etc. L'objet de la simulation, en ce cas, est donc moins le sujet psychologique de l'action que le contexte lui-même, douée de propriétés relationnelles, immédiatement lisibles, sur la valeur, l'intérêt ou le danger d'un comportement réalisé dans ce contexte.

1.2 La variété des intentions

Ces découvertes conjointes réalisées dans le champ des neurosciences cognitives ont soulevé le projet ambitieux de faire dériver la cognition sociale humaine du champ plus restreint de la simulation motrice, et de fonder, ce faisant, une « cognition sociale de la théorie motrice » qui dépendrait pour l'essentiel des découvertes opérées dans le champ de la cognition motrice (Gallese et al., 2004). Or, la possibilité même de pouvoir épuiser le champ des « cognitions sociales » dans une étude exhaustive des mécanismes sous-jacents à la simulation motrice, et à l'action en général, ne fait pas consensus.

A eux seuls, les processus élémentaires de la simulation ne suffisent pas, en effet, à rendre pleinement compte de la faculté de « lecture mentale », c'est-à-dire de cette capacité à se représenter et à attribuer à autrui des états mentaux spécifiques. Comme le soulignent dans une revue critique Pierre Jacob et Marc Jeannerod, les « systèmes-miroirs » mis en évidence chez le singe, puis chez l'homme, participent sans doute de la compréhension des actions exécutées par nos semblables, mais en un sens où l'action ainsi comprise n'a de signification que *motrice*, et qu'elle n'engage aucune interaction avec autrui (Jacob & Jeannerod, 2005).

Dans l'étude princeps de Gallese et Rizzolatti, les neurones-miroirs du singe déchargent lorsque celui-ci observe l'un de ses congénères saisir un fruit. Ici, l'activité du système-miroir du singe observateur doit effectivement lui permettre de comprendre l'intention motrice sous-jacente au geste de préhension du singe qu'il observe (« il prend ce fruit car il a faim »), mais rien de plus. A ce titre, les neurones-miroirs et les mécanismes élémentaires de simulation qu'ils sous-tendent, se borneraient à nous renseigner sur les intentions motrices de l'agent observé. Or, l'activité mentale d'un sujet ne se réduit pas à la seule formation d'intentions motrices, et les intentions motrices elles-mêmes n'épuisent pas la sphère des intentions possibles qu'un agent peut former. Selon Jacob et Jeannerod, il faudrait en réalité distinguer au moins deux types supplémentaires d'intentions, pour lesquels les neurones-miroirs demeurent muets et dont la simulation motrice ne peut rendre compte à elle seule, à savoir, l'intention privée (*prior intention*) et l'intention sociale (*social intention*).

Cette distinction n'est pas nouvelle et puise dans une typologie des intentions esquissée par John Searl (1983), puis développée ou renouvelée ces vingt dernière années dans le champ de la philosophie de l'esprit et des neurosciences (*future- and present-directed intention* : Bratman, 1987; *prospective and immediate intentions* : Mele, 1992; *distal, proximal, and motor intentions* : Pacherie 2000, 2008). Cette distinction entre niveaux intentionnels est motivée, au plan conceptuel comme empirique (voir Pacherie, 2000), par l'observation que nos actions et nos buts peuvent être ordonnés hiérarchiquement selon leur niveau d'abstraction ou selon le temps requis pour les réaliser (Hamilton & Grafton, 2006 ; v. figure 5, p.68).

Ainsi,

1. L'intention motrice, ou « intention en action » (Searle, 1983), est une intention minimale, qui, combinée à d'autres intentions du même type, peut entrer dans la composition d'une intention privée, formulée en amont de l'action elle-même. Par exemple, je peux vouloir éteindre la lumière d'une pièce suffisamment ensoleillée (intention privée) et mettre en œuvre, pour ce faire, une série de séquences motrices destinées à satisfaire cette intention préalable. L'intention motrice, en ce cas, guide le mouvement de ma main en direction de l'interrupteur que je souhaite actionner, et, ce faisant, contribue à satisfaire mon intention « privée », qui est d'éteindre la lumière.
2. Mon intention privée correspond, à ce titre, à la représentation du but que je souhaite atteindre – par la mise en œuvre d'une série d'actions basiques, elles-mêmes guidées par des intentions motrices ou « intentions en action » élémentaires. Or, si l'activité de mes neurones-miroirs peut éclairer les intentions motrices de l'agent que j'observe agir (de type « presser un interrupteur »), elle est néanmoins incapable à elle seule de me renseigner sur son intention privée, dont le contenu, général et descriptif, n'est pas directement ancré dans la situation d'action (« éteindre » ou « allumer la lumière »).

3. Enfin, toutes les actions humaines ne sont pas nécessairement dirigées vers des objets inanimés. Il importe donc d'opérer une seconde distinction entre des intentions non-sociales (qui n'entrent pas dans la composition d'un rapport avec autrui) et des intentions sociales, qui sous-tendent les actions ayant pour objectif d'affecter, de quelque façon que ce soit, le comportement d'un congénère. Or, un observateur peut-il se représenter les intentions sociales d'un agent en simulant les mouvements observés de cet agent ? L'expérience de pensée élaborée par Jacob et Jeannerod (2005) est à ce titre éloquent : Dr Jeekyll et Mr Hyde sont un seul et même individu. Le premier est un brillant médecin, reconnu pour la qualité des appendicectomies qu'il effectue. Le second est un dangereux sadique, qui réalise la même opération sur des victimes non anesthésiées. L'un et l'autre réalisent les mêmes séquences d'actions, guidées par des buts similaires (prélever l'appendice à l'aide d'un scalpel) ; pour autant, leurs intentions sociales diffèrent : Dr Jeekyll opère pour soigner tandis que Mr Hyde opère par simple plaisir sadique (figure 5, p.68). Un observateur extérieur qui simulerait les tantôt les actions du Dr Jeekyll, tantôt celles (identiques) de Mr Hyde, serait néanmoins incapable de saisir ce qui différencie le comportement du médecin de celui du criminel, à savoir leurs intentions sociales respectives (opérer ou torturer).

Les mécanismes élémentaires de la simulation motrice pourraient donc ne jouer qu'un rôle dérivé ou primitif dans la capacité proprement humaine de « mentalisation » (*mentalizing*) par laquelle nous donnons sens aux comportements de nos semblables. Cependant, il a été observé que cette compétence n'était acquise par l'enfant qu'après trois ans et demi (Wimmer & Perner, 1983). Avant cela, l'enfant ne ferait l'expérience de l'autre – de ses intentions ou de ses croyances – que sous la modalité de la simulation, c'est-à-dire en *imitant* les comportements qu'autrui manifeste au quotidien⁶. La compréhension des intentions motrices de l'autre préfigurerait ainsi la perception plus tardive d'intentions plus intégrées, privées ou sociales, dont l'accès nécessiterait, non plus une capacité robuste de simulation, mais la maîtrise d'axiomes et de règles d'inférence spécifiques, autrement dit, une *théorie naïve* de la psychologie humaine. Au total, une définition rigoureuse de notre aptitude à se représenter les intentions d'autrui requerrait donc l'élaboration d'un modèle hybride, faisant appel à la fois à un corpus de connaissance théorique et à des mécanismes de simulation (Nichols & Stich, 2003 ; Mitchell, 2005; de Lange et al., 2008). Il paraît donc crucial d'interroger les possibilités de la simulation quant à ce qu'elle peut saisir des intentions d'autrui, et de s'attacher à mieux définir les processus cognitifs en jeu dans cette aptitude, progressivement acquise, à se représenter, juger, ou raisonner à propos des états mentaux de nos semblables. Comme en témoigne par ailleurs l'intérêt croissant suscité par le champ d'investigation des « cognitions sociales » et l'orientation actuelle des études conduites dans ce domaine, il pourrait s'avérer fécond de considérer la question et les enjeux qu'elle soulève, chez le sujet sain d'une part, et à la lumière des données cliniques de la psychopathologie d'autre part, chez des individus pour lesquels cette capacité dysfonctionne.

⁶ Cette hypothèse a toutefois fait l'objet de critiques très vives. Les processus de ToM pourraient être en place à un âge bien plus précoce que ne le supposait l'étude princeps de Wimmer & Perner ; voir Onishi & Baillargeon (2005) : « Do 15-Month-Old Infants Understand False Beliefs? » ; voir également : Surian et al. (2007) : « Attribution of Beliefs by 13-Month-Old Infants ».

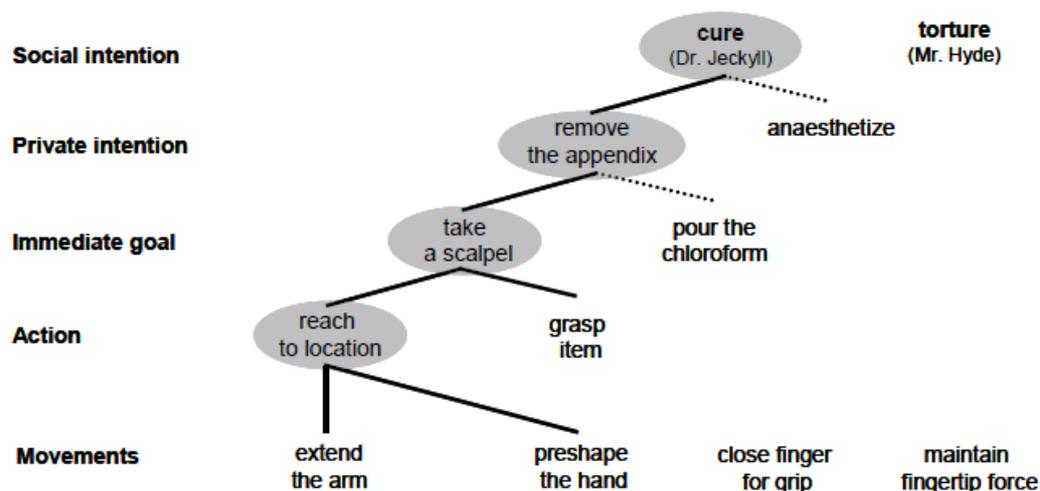


Figure 5. Organisation hiérarchique de l'action intentionnelle. Une intention sociale peut impliquer plusieurs types d'intentions privées. Chaque intention privée peut à son tour être composée de plusieurs buts immédiats, dont la réalisation requiert la mise en oeuvre d'une ou de plusieurs séquence d'actions basiques. Enfin, chaque action peut être composée de plusieurs mouvements distincts. Les buts intermédiaires désignés par les lignes en pointillés sont requis pour distinguer entre les intentions sociales "soigner" (Dr. Jeckyll) et "torturer" (Mr. Hyde). Inspiré de Hamilton et Grafton (2006), avec des exemples empruntés à Jacob et Jeannerod (2005).

2. Les capacités de représentation et d'attribution d'intentions dans la schizophrénie

La capacité de Théorie de l'Esprit joue un rôle essentiel dans nos interactions sociales : pour preuve, lorsqu'elle est défaillante, la possibilité d'interagir avec autrui s'en trouve fortement compromise, ou du moins perturbée (Frith & Corcoran, 1996; Frith & Frith, 2003). De fait, les processus de mentalisation ont fait l'objet d'une attention particulière dans le champ de la psychopathologie. Selon une théorie influente de la schizophrénie élaborée par Christopher Frith, l'ensemble des symptômes caractéristiques de cette maladie pourrait ainsi témoigner de l'incapacité des patients schizophrènes à comprendre, ou se représenter, leurs propres intentions et celles d'autrui (Frith, 1992, 2004, 2005).

L'intérêt majeur du modèle de Frith tient précisément en ce qu'il se propose de rendre compte de l'intégralité de la clinique, négative et positive, des symptômes schizophréniques dans les termes d'un défaut d'accès au plan méta-représentationnel – *i.e.* d'un défaut de représentation de ses propres intentions ou des intentions de l'autre :

1. les symptômes dits « négatifs » expriment une détérioration parfois critique du fonctionnement normal, incluant un retrait social, un émoussement affectif, une pauvreté caractérisée du langage et un déficit de comportement intentionnel. Selon Frith, ces signes comportementaux pourraient résulter d'un défaut de représentation

de l'action volontaire. Ce déficit affecterait tout comportement exigeant un effort, aussi bien que la capacité d'introspection ou la capacité d'apprécier les états mentaux des autres. Les difficultés d'adaptation sociale des patients schizophrènes, témoignant d'une perte d'intérêt pour la vie sociale en général, seraient imputables à ce déficit de lecture intentionnelle.

2. les symptômes dits « positifs » expriment un excès de fonctionnement normal, caractérisé par des aberrations du comportement, une activité délirante incohérente et des hallucinations principalement verbales. Ces aberrations du comportement pourraient résulter d'un défaut de contrôle de l'action volontaire, corrélatif d'une altération plus fondamentale de l'intention d'agir. Les idées délirantes de contrôle (syndrome d'influence), les phénomènes d'invasion de la pensée, comme les hallucinations auditives en troisième personne pourraient en effet résulter d'un trouble du contrôle de ses propres intentions (Corcoran, 2000, 2001), tandis que les idées de persécution, le délire paranoïde et le délire de référence témoigneraient de l'incapacité à inférer correctement les états mentaux des autres sur la seule base du comportement observé (Frith, 1992).

Dans le modèle influent de Frith, ce défaut d'accès au plan métareprésentationnel donnerait naissance aux symptômes, plutôt qu'il ne serait expliqué par eux. Incidemment, cette hypothèse permet la formulation de deux prédictions majeures :

i) Les sujets en rémission ne devraient pas présenter de difficultés en théorie de l'esprit ; en l'absence de symptômes, cette capacité de lecture intentionnelle devrait donc être intacte chez les patients examinés.

ii) On devrait pouvoir mettre en évidence une corrélation robuste entre les symptômes des patients testés et les déficits particuliers qu'ils présentent dans le domaine de la méta-représentation. Ces déficits, non homogènes, peuvent affecter la capacité à se représenter les croyances ou les intentions d'une personne à propos du monde (représentations de 1^{er} ordre) ou les croyances d'une personne à propos des croyances ou des intentions d'une autre personne (représentations de 2nd ordre, ou « métareprésentations »).

De nombreuses études ont été conduites afin d'évaluer la pertinence de ces prédictions. Or, les résultats obtenus sont pour le moins contradictoires : Drury et ses collaborateurs (1998) ont identifié des anomalies de la capacité d'attribution intentionnelle chez des patients schizophrènes pendant la phase aiguë de la maladie – ces anomalies disparaissant en période de rémission –, tandis qu'une étude plus récente retrouve ces perturbations en dehors même de la période aiguë, chez des patients chroniques (Herold et al., 2005).

De même, aucune association différentielle entre des déficits particuliers en « théorie de l'esprit » et des symptômes spécifiques de la schizophrénie n'a pu être établie avec certitude : certaines études indiquent que les schizophrènes déficitaires et désorganisés sont moins performants dans les tâches impliquant la maîtrise de représentations de 1^{er} ordre comme de 2nd ordre (Frith & Corcoran, 1996 ; Corcoran et al., 1995), tandis que d'autres études plus récentes ne localisent de déficits que pour les représentations de 1^{er} ordre (Mazza et al., 2001) ou de 2nd ordre seulement (Herold et al., 2003). Plus généralement, les corrélations mises en évidence entre troubles de la mentalisation et symptômes de la schizophrénie manquent de cohérence : sur le plan productif de la pathologie, plusieurs études corrélaient ce déficit avec, tantôt la sévérité des troubles de la pensée formelle non-déficitaire (Sarfati et al., 1999 ; Brunet et al., 2003), tantôt le syndrome de désorganisation

(Zalla et *al.*, 2006), tantôt le syndrome « distorsion de la réalité » (Frith & Corcoran, 1996), tandis que d'autres travaux révèlent une association significative avec le versant déficitaire de la maladie : pauvreté psychomotrice (Mazza et *al.*, 2001), affects émoussés, repli social et/ou alogie (Pickup & Frith, 2001 ; Langdon et *al.*, 2002). En règle générale, les effets de la symptomatologie sur les performances de représentation et d'attribution intentionnelles sont vagues et peu robustes, quoiqu'il ressorte, au total, que les patients schizophrènes réussissent généralement moins bien que les contrôles sains ou psychiatriques les tâches nécessitant l'appréciation des états mentaux d'autrui (Harrington et *a.l.*, 2005 ; Sprong et *al.*, 2007, pour revue).

Ces résultats incertains, souvent contradictoires, s'expliquent en partie par l'hétérogénéité des tâches employées – en termes de complexité et de matériel présenté (verbal, iconographique) – et, plus largement, par une absence de contrôle manifeste de la compétence évaluée. Ce sur quoi porte notre capacité de lecture intentionnelle, à savoir l'intention d'autrui, paraît en effet compris d'un seul bloc, alors que la notion d'intention supporte, nous l'avons vu, des niveaux de complexité variables. L'hétérogénéité des résultats mentionnés ci-dessus pourrait donc témoigner du fait que, croyant tester un même objet ou une même aptitude, les études en question évaluent en réalité cet objet à des niveaux de complexité distincts (voir Walter et *al.*, 2009, pour une critique similaire). Or, le déficit en lecture intentionnelle rapporté chez les patients schizophrènes pourrait être sélectif et très localisé, d'où l'intérêt préalable d'évaluer à *quel(s) niveau(x)* de complexité intentionnelle (intentions motrice, privée, ou sociale) ce ou ces déficits s'inscrivent, évaluation nécessairement graduelle et cohérente en cela avec une approche intégrative de la maladie.

3 La reconnaissance d'intentions : un cas particulier d'inférence probabiliste ?

Dans la schizophrénie, l'hypothèse d'un déficit en ToM est séduisante à plusieurs égards : d'abord cohérente avec l'observation clinique, elle est également étayée par un cadre théorique convaincant, dont l'un des développements majeurs est d'avoir reconnu l'*efficience causale* des états mentaux sur le comportement.

Cette hypothèse souffre néanmoins de carences conceptuelles importantes. Nous avons mentionné plus haut la nature de ces carences, qui s'expriment tant au plan théorique qu'au niveau de l'implémentation expérimentale des hypothèses. En effet, il ne suffit pas de présupposer que les états mentaux *causent* le comportement pour justifier une interprétation « mentaliste » du comportement ; il faut encore pouvoir disposer d'une typologie précise de ces états, et du rapport de causalité que ces états entretiennent avec le comportement observé. Une étude rigoureuse des capacités de lecture intentionnelle ne peut donc faire l'économie d'une évaluation graduelle de la notion d'intention, conduite à des niveaux de complexité variables – moteur (lecture des sous-buts de l'action), privé (représentation du but préalable à l'action), mais également social (compréhension de l'action dirigée vers autrui).

Nous avons vu que l'hétérogénéité des résultats observés dans la pathologie rendait nécessaire ce travail de « décomposition ». Nous verrons que ce travail est également d'une importance cruciale pour le débat qui oppose les simulationnistes aux partisans de la

théorie-théorie, dont la stricte opposition se trouve en effet résolue dès lors que *i*) on ne néglige aucun des *niveaux* de la typologie mentale concernée, et *ii*) que l'on précise à quel niveau de cette typologie on se situe. Les processus de simulation (de type « miroirs ») et les concepts mentaux primitifs postulés par la théorie-théorie contribuent sans doute ensemble aux facultés de mentalisation : reste néanmoins à déterminer le poids de leur contribution respective, contribution qui pourrait dépendre du type d'état mental étudié, et de sa complexité relative. Par nécessité, nous plaiderons donc en faveur d'un modèle *hybride* des capacités de reconnaissance intentionnelle, mobilisant tantôt des processus perceptuo-moteurs de bas-niveau (de type « miroirs »), tantôt un ensemble de connaissances *a priori* permettant à l'observateur d'inférer, depuis l'expérience, l'ensemble des règles qui sous-tend le comportement observé.

3.1 Vraisemblance perceptive et informations *a priori*

Pour tester la pertinence de ce modèle, nous avons élaboré plusieurs protocoles expérimentaux destinés à tester, individuellement, chacun des niveaux de la typologie intentionnelle mentionnés ci-dessus. Pour rappel, cette typologie distingue deux grandes classes d'intentions : motrice et privée, à composante sociale (lorsque l'action est à *propos* d'autrui) ou non-sociale (lorsque l'action est à *propos* d'un objet). Une fois définie la granularité de cette typologie, la question que nous avons adressée était double :

1. toutes ces intentions (motrice et privée, sociale ou non-sociale) partagent-elles les mêmes mécanismes de reconnaissance ? Peut-on les distinguer au regard du type d'information qui contribue à leur reconnaissance ?
2. le déficit (apparent) des patients schizophrènes en matière de mentalisation est-il spécifique d'une intention particulière, ou d'un mécanisme associé à l'un ou l'autre niveau de cette typologie ?

Pour répondre à ces deux questions, nous avons formulé l'hypothèse de travail suivante : reconnaître une intention implique de traiter au moins deux types d'information, 1) des informations de nature *perceptive*, disponibles sur l'action en cours de réalisation, et 2) des informations *a priori*, relatives aux connaissances préalables que je possède sur l'action observée. Ces connaissances *a priori* sont naturellement susceptibles d'interagir avec ce que nous percevons de la scène d'action. En contexte d'incertitude perceptive par exemple, ces *a priori* favoriseront le poids de certaines prédictions, en augmentant la probabilité de certaines conclusions, au détriment d'autres. Lorsque l'exercice consiste à déterminer la cause la plus probable de l'action observée, cette interaction entre *a priori* et données perceptives peut se traduire par une augmentation de la probabilité de certaines causes – de certaines intentions – au détriment de leurs concurrentes. Nous ne discuterons pas la nature de ces connaissances *a priori*, qui a fait l'objet de vives controverses⁷ ; seule nous intéresse ici la fonction que ces connaissances remplissent dans l'économie générale du système : en contexte d'incertitude (lorsque l'action est incomplète, bruitée ou ambiguë), ces connaissances *a priori* auront pour effet de favoriser, en l'augmentant, la probabilité que

⁷ La fonction que remplissent ces *a priori* pourrait être prise en charge par une classe particulière de neurones – appelés « logically related neurons » – sensibles à la prédictibilité des chaînes d'actions élémentaires (voir Iacoboni et al., 2005). Cette classe de neurones, dont il reste encore à démontrer l'existence, plaide naturellement en faveur des théories simulationnistes. Il existe néanmoins des alternatives, lesquelles militent plutôt en faveur d'un modèle « hybride » des facultés de mentalisation : dans le cadre du formalisme Bayésien, il est ainsi possible de concevoir les informations *a priori* du sujet comme des « théories de niveaux supérieurs » (Baker et al., 2006), susceptibles de réduire l'« erreur de prédiction » que génère l'information perceptive lorsqu'elle est ambiguë, bruitée ou incomplète (Kilner et al., 2007).

certaines hypothèse soient vraies, ou que certaines causes soient réelles. Lorsqu'il s'agit de reconnaître l'intention d'autrui, le rôle de ces *a priori* peut être crucial : l'intention est un état inobservable et le comportement est souvent tronqué, ou ambigu. Le succès de l'inférence dépendra donc étroitement de ce que l'on sait *a priori* des causes (des intentions) qui, en général, motivent l'exécution de ce comportement.

S'il est aujourd'hui admis que ces deux sources d'information, perceptive et *a priori*, interagissent au cours du processus de reconnaissance, il reste encore à déterminer la nature exacte de cette contribution. Cette question fait naturellement écho à celle que nous adressions au paragraphe précédent. Les mécanismes de la simulation traitent essentiellement des informations de nature *perceptive*, liées au mouvement biologique, tandis que les concepts primitifs postulés par la « théorie-théorie » réfèrent à un ensemble de connaissances *préalables* à l'expérience perceptive du mouvement. De fait, déterminer le rôle que jouent informations perceptives et *a priori* au cours du processus de reconnaissance intentionnelle revient à préciser la contribution respective de ces deux grandes classes de mécanismes (simulation ou concepts primitifs) dans le succès de ce processus. Nous avons formulé l'hypothèse que la nature de cette contribution dépendait en réalité du *type* d'intention présenté : informations *a priori* et information perceptives devraient naturellement jouer un rôle distinct selon le niveau de la typologie étudié (intention motrice ou privée, sociale ou non-sociale).

3.2 L'apport du formalisme Bayésien

Pour tester cette hypothèse, les neurosciences computationnelles nous ont été d'un secours précieux. Il existe en effet un outil mathématique particulièrement adapté pour modéliser le poids respectif de ces deux sources d'information – perceptives et *a priori* –, ainsi que leur interaction au cours du temps. Cet outil, représenté sous la forme d'un théorème appelé « théorème de Bayes », a déjà fait l'objet d'une attention particulière dans des domaines aussi variés que l'apprentissage (Knill & Pouget, 2004), le contrôle sensorimoteur (Körding & Wolpert), ou la perception (Mamassian & Goutcher, 2001). Dans le cadre du formalisme Bayésien, l'acte de perception, par exemple, peut être modélisé comme un processus d'inférence probabiliste, en partie non-conscient. Cette inférence combine deux paramètres : une probabilité *a priori* (*i.e.* la probabilité de chaque état possible de la scène avant que ne survienne le stimulus), et une *vraisemblance*, relatives aux données perceptives de la scène (*i.e.* la probabilité du stimulus étant donné chaque état possible de la scène). Cette approche peut naturellement être généralisée aux processus cognitifs de plus haut-niveau (Tenenbaum et al., 2007 ; Baker et al., 2006), et nous avons fait l'hypothèse que les processus engagés dans la reconnaissance d'intentions devaient eux aussi obéir à une logique Bayésienne. Dans ce cadre, l'inférence intentionnelle devrait alors pouvoir être modélisée comme le produit d'une interaction entre des probabilités *a priori* (la probabilité qu'une intention cause le comportement observé) et des données sensorielles, ayant une certaine vraisemblance (les informations perceptives que le sujet accumule sur la scène d'action).

Quatre protocoles expérimentaux ont été élaborés pour tester cette hypothèse. Chaque protocole présentait des séquences d'actions filmées, guidées par des intentions de nature motrice ou privée (sociale ou non-sociale), que les participants devaient inférer. Dans chacune de ces expériences, et conformément à notre « assomption Bayésienne », nous avons donc fait varier

i) la vraisemblance des informations visuelles, disponibles sur la scène d'action présentée (pour ce faire, nous manipulons la durée de séquences d'actions filmées ; 'COVERT blocks', figure 6, page suivante),

ii) les connaissances *a priori* dont disposait le sujet sur les types d'intentions à reconnaître (pour ce faire, nous augmentons la probabilité de certaines intentions, au détriment d'autres ; 'OVERT blocks, figure 6).

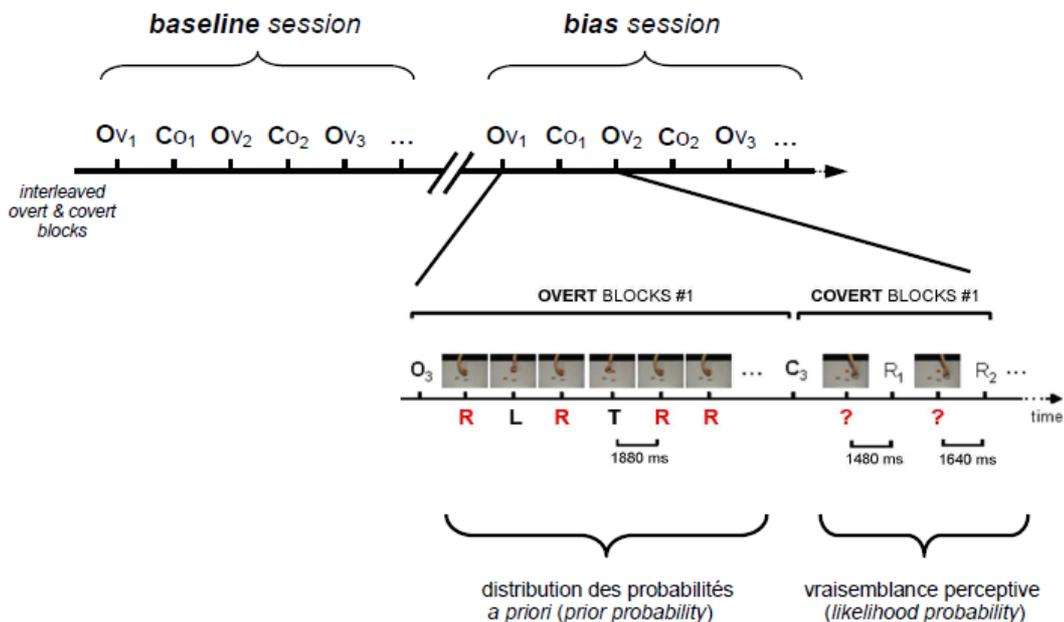


Figure 6. Design expérimental : exemples de séquences expérimentales typiques (un block "OVERT" suivi par un block "COVERT") utilisées dans les sessions BASELINE et BIAS. Overt blocks: videos pour lesquelles la quantité d'information visuelle est élevée et constante. Covert blocks: videos pour lesquelles la quantité d'information visuelle varie sur trois niveaux (LOW, MODERATE, HIGH). Dans les 4 expériences présentées, toutes les intentions ont la même probabilité d'être réalisées, à l'exception des "OVERT blocks" de la session BIAS où une intention particulière a une probabilité d'occurrence plus importante que les deux autres (ici, l'action "tourner" (rotate), en rouge). Labels: Ov: Overt blocks; Co: Covert blocks; L: "lift" action; R: "rotate" action; T: "transport" action.

Nous avons ensuite comparé les performances des participants entre ces conditions dites « biaisées », où certaines intentions étaient plus probables que d'autres ('bias session'), et des conditions « contrôles » ('baseline session') où toutes les intentions présentées avaient la même probabilité d'être réalisées. Nos prédictions étaient les suivantes : si l'inférence intentionnelle obéit à une logique Bayésienne, alors

1. dans les conditions où l'action est présentée avec une quantité d'information visuelle peu élevée, les sujets devraient tabler davantage sur leurs connaissances *a priori* – c'est-à-dire avoir tendance à répondre préférentiellement en direction de l'intention qu'ils estiment la plus probable ;
2. en revanche, lorsque l'action est présentée avec une quantité d'information visuelle élevée, les sujets devraient être moins sensibles à ce biais et présenter une dépendance accrue aux informations perceptives.

Notre seconde prédiction concernait la *qualité* de l'interaction entre ces deux sources d'informations, dont nous avons estimé qu'elle devait dépendre du *type* d'intention testé :

i) Pour la reconnaissance d'intentions **motrices** simples, les sujets devraient présenter une dépendance accrue aux informations perceptives. Cette prédiction est motivée par la nature *pragmatique* du contenu de l'intention motrice : il existe en effet une correspondance directe entre « l'intention de lever un verre » et l'acte de « lever » ce verre. En ce cas, *percevoir* l'acte lui-même devrait suffire à déterminer la nature de l'intention sous-jacente.

ii) Pour les intentions **privées** (sociale ou non-sociale) en revanche, les sujets devraient présenter une dépendance accrue aux informations *a priori*. Le contenu de l'intention privée (« avoir l'intention d'éteindre la lumière ») est en effet général et descriptif, et il est en partie détaché de la situation d'action : l'information perceptive sur le mouvement exécuté (« presser l'interrupteur » par exemple) ne suffit donc pas à déterminer la nature de l'intention privée, car cette information sous-détermine l'ensemble des intentions privées congruentes avec ce mouvement (l'acte « presser l'interrupteur » est en effet compatible aussi bien avec l'intention privée « éteindre » ou « allumer » la lumière). De fait, la reconnaissance d'une intention de nature privée s'opère dans un contexte d'incertitude perceptive qui devrait encourager les participants à se « méfier » de ce qu'ils observent et à se fier davantage à leurs attentes *a priori*.

iii) Lorsque les intentions motrices ou privées présentent une composante **sociale** (elles sont à *propos* d'autrui), cette dépendance aux informations *a priori* devrait s'accroître. La structure de l'interaction sociale (collaborative ou compétitive) répond en effet à des attentes particulières, souvent irrépressibles (la réputation des agents engagés dans l'interaction, par exemple ; Fu & Lee, 2007 ; Frith & Frith, 2006, pour revue). Nous estimons que le poids de ces attentes *a priori* devraient influencer massivement les prédictions du sujet, jusqu'à l'emporter sur la vraisemblance perceptive des séquences d'actions présentées.

iv) Chez les patients schizophrènes, enfin, nous nous attendions à ce que le déficit observé pour l'un ou l'autre niveau de la hiérarchie intentionnelle, s'exprime sous la forme d'une interaction inappropriée entre informations *a priori* et vraisemblance perceptive. Cette interaction anormale pourrait traduire une confiance excessive accordée à une source d'information au détriment de l'autre. La nature de l'information privilégiée devrait en outre dépendre du profil symptomatologique du patient : le profil de mentalisation des patients « productifs » pourrait en effet se caractériser par une tendance à accorder un crédit excessif aux informations endogènes et auto-générées (Firth, 1992), tandis que les patients « négatifs » inclineraient à se concentrer sur les informations exogènes et directement observables plutôt que sur leurs propres expériences internes (Taylor, 1994).

Nous avons donc systématiquement évalués, pour chacun des patients testés, la sévérité des dimensions productive et négative *via* les échelles de cotation clinique standard, SANS (Andreasen, 1983) et SAPS (Andreasen, 1984).

4 Protocoles expérimentaux : intention motrice, intention superordonnée, intentions sociales

Quatre groupes de 30 sujets témoins et 4 groupes de 20 patients schizophrènes ont accepté de participer à l'étude, divisée en quatre expériences distinctes (figure 7, p.80).

Dans la première expérience, les participants avaient pour instruction de reconnaître les *intentions motrices* (lever, tourner ou tirer) d'un acteur manipulant des cubes disposés devant lui (**A**). Dans la seconde expérience, il était cette fois demandé aux participants de reconnaître le *but général* d'une séquence de construction : la séquence consistait à nouveau en trois actions simples (lever, tirer ou tourner), successivement réalisées sur trois cubes identiques à ceux manipulés dans la première étude (**B**). La forme construite représentait ici *l'intention privée* de l'acteur, que les participants devaient inférer. Dans la mesure où toutes les intentions privées ne sont pas nécessairement le résultat d'une *séquence* d'actions articulées, nous avons choisi ici de renommer l'intention dite « privée » en intention « **superordonnée** », *i.e.* une intention dont le contenu (« éteindre la lumière ») super-ordonne la séquence d'actions réalisées (« tendre le bras », « éteindre l'interrupteur », etc.).

La troisième (**C**) et la quatrième expérience (**D**) mettaient en scène deux acteurs, engagés dans un jeu de compétition/collaboration. Le jeu s'organisait autour d'un damier composé de cubes, que les acteurs pouvaient tourner ou déplacer latéralement, certaines configurations encourageant la collaboration, d'autres la compétition. La tâche des participants consistait à inférer le type de stratégie mise en œuvre par le second joueur, étant donné le type de stratégie préalablement adopté par le premier joueur. Dans cette condition à composante sociale, nous manipulions également des intentions *motrices* (tourner ou déplacer les cubes) et *superordonnées* (la forme construite sur le damier), modulées par les intentions préalables du joueur précédent (*i.e.* sa « réputation »).

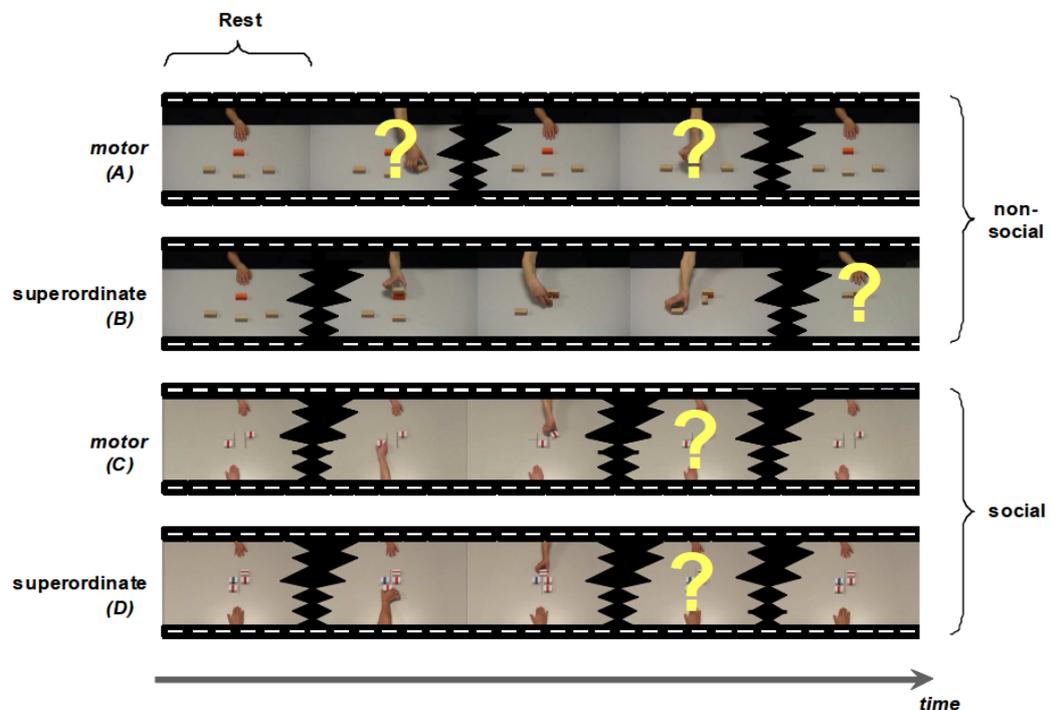


Figure 7. Séquences d'actions typiques : (A) condition motrice (non-sociale); (B) condition superordonnée (non-sociale); (C) condition motrice sociale; (D) condition superordonnée sociale.

Facteur information visuelle (pré-test)

Dans les quatre expériences, les vidéos d'action étaient présentées avec 4 niveaux d'informations visuelle distincts : très élevée, élevée, modérée, et basse {1880, 1640, 1560, 1480 ms, respectivement}. De manière à contrôler la dynamique temporelle des mouvements présentés et l'homogénéité des vidéos entre elles, toutes les séquences filmées ont été égalisées en nombre d'images (*frames*) par seconde – pour chaque niveau d'information visuelle – et pré-testées sur deux groupes de 12 sujets (voir ci-dessous, Chambon et *al.*, Appendix). Les corrections ont été effectuées en appliquant un algorithme d'interpolation (SMO_intp, Adobe Premiere®) aux séquences trop lentes ou trop rapides. Pour les séquences trop lentes, la procédure consistait à *remplacer* deux images contiguës par une image qui représentait la moyenne calculée de ces deux images ; pour les

séquences trop rapides, la procédure consistait à *intercaler* entre ces deux images, leur image moyenne.

Facteur probabilité

Deux facteurs étaient manipulés durant l'expérience : le niveau d'information visuelle des séquences d'actions d'une part (cf. *supra*), et, d'autre part, la probabilité que certaines intentions soient réalisées, au détriment d'autres :

1. la probabilité que l'acteur réalise l'action « tourner », par exemple, au détriment des deux autres actions possibles (« soulever », « tirer ») [intention motrice]
2. la probabilité qu'une forme particulière (but général de la séquence) soit construite, au détriment des deux autres [intention superordonnée]
3. la probabilité que le second joueur adopte une stratégie « coup-pour-coup », i.e. joue un coup « en miroir » de ce qu'a fait le premier joueur au tour précédent (e.g. coopère si le premier joueur a coopéré, fait défaut si le premier joueur a fait défaut). Cette stratégie apparaissait au détriment des stratégies alternatives, de type « égoïste » (le joueur fait toujours défaut), « altruiste » (le joueur coopère toujours), ou « random » (le joueur n'a pas de stratégie particulière). [intentions sociales]

Pour chacune des 4 conditions intentionnelles, nous avons comparé entre elles les performances des participants pour la reconnaissance des intentions biaisées (probabilité élevée) vs. non-biaisées (probabilité basse), et cela, pour les trois niveaux d'information visuelle manipulée durant la tâche (élevée, modérée, et basse). Le premier article présente les résultats obtenus chez les sujets sains non-psychiatriques, pour les 4 études présentées ci-dessus. Le second papier présente les résultats des patients schizophrènes, divisés en 4 groupes que nous avons appariés sur les scores obtenus à la SANS (Andreasen, 1983) et à la SAPS (1984). Les performances des patients ont été directement comparées aux performances des sujets témoins de la première étude.

Etude chez le sujet sain - What are they up to? The role of sensory evidence and prior knowledge in action understanding

Valerian Chambon, Philippe Domenech, Elisabeth Pacherie, Etienne Koechlin, Pierre Baraduc and Chloé Farrer

(soumis)

Abstract

Explaining or predicting the behaviour of our conspecifics requires the ability to infer the intentions that motivate it. Such inferences are assumed to rely on two types of information: (1) the sensory information conveyed by movement kinematics and (2) the observer's prior expectations – acquired from past experience or derived from prior knowledge. However, the respective contribution of these two sources of information is still controversial. This controversy could stem in part from the fact that “intention” is an umbrella term that may embrace various sub-types each being assigned different scopes and targets.

We hypothesized that variations in the scope and target of intentions may account for variations in the contribution of visual kinematics and prior knowledge to the intention inference process. To test this hypothesis 4 behavioural experiments were conducted in which participants were instructed to identify different types of intention: motor intentions (i.e. simple goal of a motor act), superordinate intentions (i.e. general goal of a sequence of motor acts), or social intentions (i.e. intentions accomplished in a context of reciprocal interaction). For each of the above-mentioned intentions, we also varied (1) the amount of visual information available from the action scene and (2) participant's prior expectations reflecting probability of achievement of different intentions. First, we showed that intentional judgments depend on a consistent interaction between visual information and participants' prior expectations. Crucially, this interaction varies according to the type of intention to be inferred, with participant's priors exerting a greater effect on the judgment of social and superordinate intentions, at the expense of sensory evidence. The results are discussed by appealing to the specific properties of each type of intention considered and further interpreted in the light of a hierarchical model of action representation.

Keywords: Intention recognition; Sensory evidence; Prior knowledge; Bayesian inference; Social interaction.

1. Introduction

1.1. Intentional inference: perceptual information and top-down prior knowledge

Explaining or predicting the behaviour of our conspecifics requires the ability to properly appreciate the causes that motivate it. As the observed behaviour is self-initiated and directed at a specific goal, these causes are hidden rather than visible: intentions – like beliefs or desires – are unobservable states. Thus, it has long been a matter of speculation how one may infer them from mere observation, that is, from patterns of *visible* behaviour alone.

One decade ago, this question got a resounding response with the discovery, in the monkey and then in humans, of a system of cerebral regions endowed with 'mirror' properties. Neurons from these regions discharge *both when* a subject performs a given action and sees the same action performed by a third party (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996; Grafton, Arbib, Fadiga, & Rizzolatti, 1996). This process of "motor equivalence" speaks convincingly in favor of a direct relationship between action observation and action understanding. During action observation, visual information conveyed by movement kinematics automatically maps onto the observer's own motor system, then activating representations similar to those the subject would have generated during active action (Rizzolatti & Craighero, 2004). The activation of these 'shared' representations – activation that is similar to a process of internal replication, known as 'simulation' – would provide a direct access to the intentions of the agent, thus allowing prediction to be made about her current behaviour (Gallese & Goldman, 1998; Grèzes & Decéty, 2001; Rizzolatti, Fogassi, & Gallese, 2001; Jeannerod, 2001; Blakemore & Decéty, 2001).

However, a problem arises with this model. In the "mirror" account of action understanding, visual information from the observed movement is supposed to be 'directly' transformed into knowledge about the action goal by inverting the model we generally use to perform an action (from current visual kinematics to action plans congruent with

this kinematics). Thus, a one-to-one mapping is implicitly assumed between intention and action: one particular action denotes one particular intention that can be inferred once the corresponding visual kinematics is properly processed. But as a rule, the relation between mental states and observable behaviour is more complex than a one-to-one correspondence (Oztop, Wolpert, & Kawato, 2005; Jacob, 2008; see also Becchio, Sartori, Bulgheroni, & Castiello, 2008). Besides, information conveyed by kinematic movement is often noisy, ambiguous, or incomplete. As a result, visual information generally under-constrains the space of candidate causes (i.e. the many competing intentions) that are logically consistent with what is observed (Kilner, Friston, & Frith, 2007a, 2007b; Csibra & Gergely, 2007).

One way to solve this problem is to assume that this space of possible intentions is further constrained by the observer's prior expectations. These expectations are derived from a prior knowledge that may originate from the past experience of the viewer (through expertise: Sebanz & Shiffrar, 2006; Aglioti, Cesari, Romani, & Urgesi, 2008, or learning of statistical regularities: Gold & Shadlen, 2002), from her intuitive theories (Carey, 1985; Gopnik & Wellman, 1994), or reputational knowledge (King-Casas et al., 2005; Singer et al., 2006), as well as from contextual information surrounding the action scene (Sebanz et al., 2006; Aglioti et al., 2008). This prior knowledge has been demonstrated crucial to account for the robustness of our everyday inferences (see Frith & Frith, 2006). Indeed, by exerting top-down influences onto the observer's shared representations, it makes possible inductive inference about the agent's intentions, even in case of noisy signals or incomplete data (Baker, Tenenbaum, & Saxe, 2006, 2007; Griffiths, Kemp, & Tenenbaum, 2008).

However, although most authors agree that prior knowledge and perceptual information both contribute to the process of inferring intentions, the precise contribution of each type of information remains controversial (see Gallese, Keysers, & Rizzolatti, 2004; Jacob & Jeannerod, 2005; Saxe, 2005; Goldman & Sebanz, 2005; Mitchell, 2005; Nichols & Stich; Keysers & Gazzola, 2007; deLange, Spronk, Willems, Toni, & Bekkering, 2008). The controversy appears to stem in part from the fact that "intention" is an umbrella term used in the empirical and philosophical literature to refer to representations of actions that can differ in both their content and format, as well as in their temporal properties and in the role they play in the guidance of actions (Searle, 1983; Bratman, 1987; Mele, 1992; Pacherie 2000, 2008; Jeannerod & Pacherie, 2004). Intentions can therefore be distinguished into several sub-types according to one or several of these factors. In the present study, we propose a typology of intentions and hypothesize that this typology might be a key element in understanding how perceptual information and prior knowledge contribute to the process by which intentions are inferred.

1.2. Varieties of intentions

The intentional typology we present below is primarily motivated by the necessity to take into account two dimensions of variation in the content of intentions that may make an important difference to the processes involved in their inference. The first dimension of variation concerns the scope of the intention; i.e., the more or less complex nature of its goal. Here we can draw a distinction between:

(i) *Motor intentions*: They are intentions directed at simple motor goals, i.e. goals that can be realized by basic actions (such as lifting an arm, pressing a button, reaching for an object, etc.). Motor intentions are sensorimotor representations where the goal is represented directly in terms of the motor commands needed to achieve it. The relation between motor intention and motor act is thus a one-one when that act is successfully completed.

(ii) *Superordinate intentions*: They are intentions directed at somewhat complex or general goals, the achievement of which typically involves the completion of a number of subgoals or substeps. Depending on the complexity of the general goal, these subgoals may themselves be decomposed into further subgoals, ultimately reaching the level of basic actions. The achievement of a superordinate intention will thus require the execution of a combination of basic actions each guided by a corresponding motor intention. Different combinations of motor acts can be used to accomplish the same general goal and, conversely, a same motor act (or even series of motor acts) can be part of combinations aimed at different general goals. As a result, when visual information on motor acts is noisy or incomplete, recognizing the superordinate intention driving these acts may become a challenge.

The second dimension of variation we were interested in concerns the target of the intentions. Neither motor nor superordinate intentions are necessarily directed at inanimate objects. They may also target a third party or be achieved in a context of social interaction (Searle, 1995; Blakemore & Frith, 2004; Sebanz, Bekkering, & Knoblich, 2006; Frith & Frith, 2007; Ciaramidaro et al., 2007). Intentional contents are thus also modulated by the relational structure in which an action takes place. We called intentions directed at an object, *non-social* intentions, and intentions directed at a third party, *social* intentions.

By combining these two dimensions, we obtain the following typology: i) *non-social* motor intention, ii) *non-social* superordinate intention, iii) *social* motor intention, and iv) *social* superordinate intention.

Owing to their different scopes and targets, motor and superordinate, social and non-social intentions are naturally assigned different functional roles, different types of content and different temporal scales. The present study aims at investigating whether these functional differences would be reflected in the respective contribution of perceptual information and prior expectations to intentional judgments.

1.3. Overview of the present study

To test this hypothesis, we conducted four experiments in which participants were requested to infer one intention underlying an action scene. Each experiment involved one type of intention with a specific scope (motor vs. superordinate) and a specific target (social vs. non-social). Interactions between prior expectations and visual information were examined within a Bayesian probabilistic framework. This conceptual framework is particularly well-suited to account for how accurate predictions on hidden world states are made in situations where available sensory information does not sufficiently constrain the number of potential solutions (Baker et al., 2006; Körding & Wolpert, 2006; Griffiths et al., 2008). Before the onset of an action sequence, each of the agent's possible intentions is first assigned a certain 'level of belief', termed *a priori probability* (the probability that intention X is the real cause of the observed behaviour estimated from past experiences). Then the observer progressively gathers sensory information (visual input) as the action sequence is disclosed and both sources of information (sensory and *a priori*) are combined to infer the intention that motivates the observed behaviour. Thus, the process by which intentions are inferred is considered as reflecting a trade-off between the sensory information and the prior probability of each candidate intention (Baker et al., 2007). Finally, the chosen intention is that which maximizes the *posterior probability* value, i.e. the probability that intention X is true *given* what is observed.

In the present study, these two terms – *a priori* probability and sensory information – were manipulated using a two-steps procedure:

(i) Prior expectations the participants had about the agent's possible intentions were manipulated by increasing the *a priori* probability that one intention (termed biased or *preferred intention*) occurs, at the expense of other intentions (non-biased or *non-preferred intentions*) with the same scope and target.

(ii) *Sensory* information available from an action scene was then manipulated in a second step by modulating the degree of completeness (i.e. the duration) of the action sequences, resulting in actions scenes with varying amount of visual information.

We first predicted that judgements on intentions should follow the general principles of Bayesian inference. Specifically, we expected that the amount of visual information would interact with participants' prior expectations such that the lower the reliability of the external visual input, the more participants' responses should depend upon their own internal expectations. That is, they should respond more frequently in the direction of the bias (i.e. the preferred intention) towards which they were assigned. And vice versa, the higher the amount of visual information, the less the participants should rely on their prior expectations.

Second, we predicted that the shape of the interaction between these two sources of information should be a function of the *type* of intention, depending on both its scope and target. Along the dimension of the scope, we hypothesised that participants' judgement about motor intentions should primarily rely on sensory information available from the action scene. This prediction is motivated by the pragmatic content of the motor intention: "grasping a glass of water" directly denotes the corresponding intention of "grasping that glass". In this case, perceiving the action itself – i.e. processing the associated visual kinematics – is enough to successfully determine the nature of the underlying intention (Rizzolatti *et al.*, 2001, 2004). On the other hand, we expected performance in judging superordinate intentions to be significantly influenced by participants' prior expectations. As already mentioned, the same sequence of motor acts can be part of combinations aimed at different general goals or superordinate intentions. In this specific case, sensory information carried by kinematic movement is not sufficient to infer the corresponding intention, as it under-constrains the set of candidate intentions congruent with this movement (Jacob & Jeannerod, 2005; Kilner *et al.*, 2007a, 2007b; Jacob, 2008). We consequently predicted that this perceptual uncertainty should encourage participants to « mistrust » what they observe and, hence, to rely more on their prior expectations.

Along the dimension of target, finally, we expected participants' reliance on their prior expectations to increase when motor and superordinate intentions are directed at another agent. The structure of social interaction meets indeed particular, often irrepressible, expectations, such as those provided by reputational knowledge (King-Casas *et al.*, 2005; Singer *et al.*, 2006; Fu & Lee, 2007). Indeed, knowledge about individual's reputation has been robustly demonstrated to have a strong impact on predicting how the observed agent will behave (Frith & Frith, 2006). In line with other recent suggestions, we thus hypothesised that the weight of these *a priori* expectations should increase when the observed action fits into a context of social interaction.

2. Methods and results

The study consisted of four experiments, each combining one type of target (social vs. non-social) with one kind of scope (motor vs. superordinate). All four experiments had the same design with two distinct sessions (baseline and bias session) that differed in whether participants were, or were not, biased towards recognising one particular intention. In the

baseline session, no bias was assigned – all intentions thus had equal probabilities –, whereas, in the bias session, one particular intention (preferred intention) was more likely to be achieved than the other ones (non-preferred intentions). Both sessions consisted of interleaved blocks (overt and covert blocks) that differed in the amount of visual information that was given to the subjects. Overt blocks were characterized by a very high amount of visual information that allowed participants to easily recognize the different intentions and therefore to integrate the effective probability distribution of the different intentions associated with each session. In the covert blocks, however, action sequences carried varying and lower amounts of visual information. This design allowed us to assess how the bias (i.e. the participants' prior expectations) interacts with the amount of visual information during an intentional inference task.

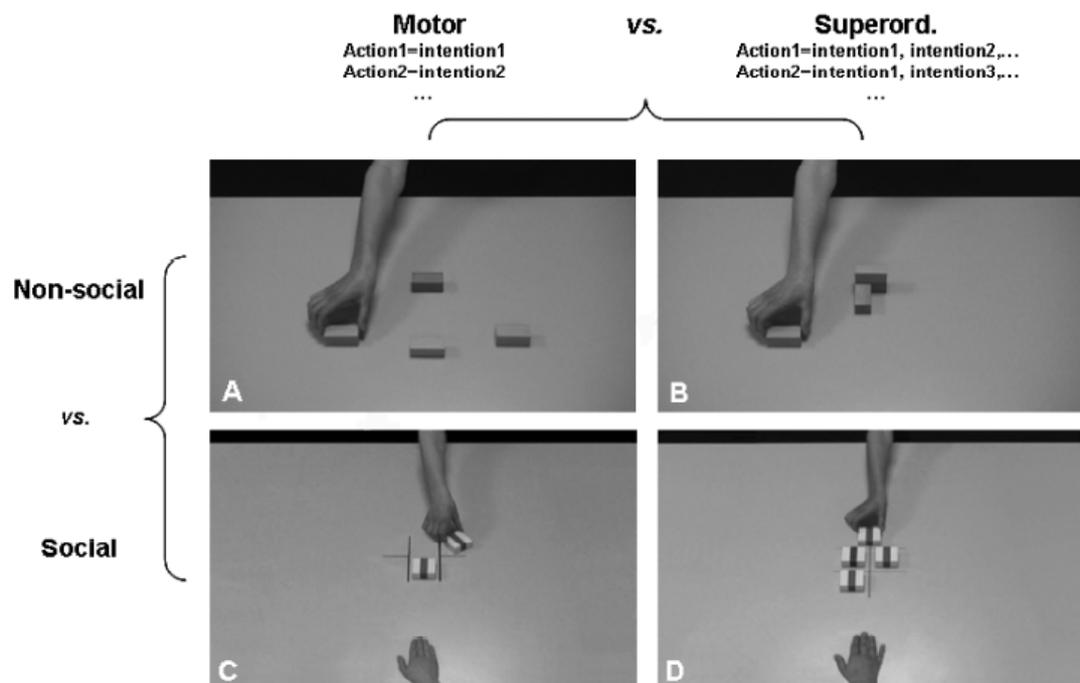


Fig. 1. Examples of stimuli for the motor non-social intention experiment (A); the superordinate non-social intention experiment (B); the motor social intention experiment (C) and the superordinate social intention experiment (D).

2.1. Non social experiments

In the first experiment, participants were instructed to infer the *motor intention* (to lift, to rotate or to transport) of an actor manipulating meaningless objects (fig. 1, **A**). In the second experiment, participants were instructed to infer the *superordinate intention* (i.e. the general goal) underlying a sequence of motor acts (fig. 1, **B**).

2.1.1. Experiment 1: non-social motor intention

2.1.1.1. Methods

Participants

30 healthy participants (15 males, 15 females, mean age = 35.13, S.D. = 9.33 and laterality score mean = 0.88, S.D. = 0.31; Oldfield, 1971) participated in this experiment. They all reported normal or corrected-to-normal visual acuity. Participants received 10 euros for taking part in the study.

Stimuli

Visual stimuli presented were incomplete movies representing an actor's hand performing a simple manipulation of a meaningless object. The duration of the video sequences was varied on 4 distinct levels, ranging from 1480 ms to 1880 ms after movement onset. Each movie was characterized by one motor intention (to transport {Im1}, to rotate {Im2}, or to lift {Im3} a cube) that participants were instructed to infer. The objects used were three rectangular cubes of similar size (3cm x 6cm) and orientation, placed at equal distance (16,8 cm) from the starting position of the actor's hand (figure 1, A). Actions were performed on one of the cubes in the direction of the centre of the scene. All the movies were performed by a single actor and only featured her naked arm. Each action was performed equally often with all objects (cube 1, 2, or 3). The movies were presented on a computer monitor (IIYAMA® 19") at a distance of 60 cm from the participant.

Movies were equalized for temporal homogeneity in such a way that they all were composed of the same number of frames (see Appendix, part A). Furthermore, each trial was unique; all the movies were presented only once to prevent any influence of memorized kinematics parameters on participants' performance in the experiment.

Procedure

Each trial started with the onset of an incomplete movie. As soon as the sequence ended, three letters, corresponding to the first letter of the three possible intentions (T for 'transport', L for 'lift', or R for 'rotate') appeared on a black screen. Participants were requested to respond by pressing, as quickly and as accurately as possible, one of the three keyboard buttons corresponding to the three possible intentions. Once a response was given, the next trial started.

The design was composed of two experimental sessions and a two-steps procedure was therefore employed:

(1) **Baseline session.** This session was characterized by a flat (unbiased) prior probability distribution with all motor intentions having the same probability ($p\{Im1\}=p\{Im2\}=p\{Im3\}$). Trials were organized into blocks in which movies consisted of a constant and very high amount of visual information (overt blocks), alternating with blocks in which movies consisted of varying and lower amounts of visual information (covert blocks).

Each experimental sequence (one overt block followed by one covert block) was repeated 9 times over the session (see figure 2, A).

1.1: *overt* blocks of movies with a constant amount of visual information.

In these blocks, the videos were first presented with a very high amount of visual information (1880 ms) (see Appendix, part B, for the selection of that amount). Each motor intention ($\{Im1\}$, $\{Im2\}$, $\{Im3\}$) was repeated 6 times within a block, resulting in 18 trials for the whole session. Each participant performed the trials in a different pseudo-random order.

1.2: *covert* blocks of movies with varying amounts of visual information.

Movies were partitioned into sequences with low, moderate and high amounts of information (respective durations: $Inf1 = 1480$ ms, $Inf2 = 1560$ ms, and $Inf3 = 1640$ ms; see Appendix, part B, for the selection of these amounts). Each covert block involved two different amounts of information ($Inf[1,2]$, $Inf[1,3]$, or $Inf[2,3]$) presented in a pseudo-random order. In these blocks, the probability of each motor intention was held equal across the block ($p\{Im1\}=p\{Im2\}=p\{Im3\}$).

(2) **Bias session.** In a second session, which also consisted of interleaved blocks, participants were biased towards one intention (preferred intention) to the detriment of the others (non-preferred intentions). In the overt blocks, the preferred intention had a higher probability to be achieved compared to the other ones (e.g. $p\{Im3\} > p\{Im1\} = p\{Im2\}$). This procedure allowed the assignment of a specific bias (i.e. the preferred intention) to the participants. Each experimental sequence was repeated 9 times over the session. The bias was pseudo-randomly assigned so that each motor intention was equally biased across participants.

2.1: *overt* blocks of movies with unequal probabilities.

In these blocks, one motor intention (preferred intention) was most frequently shown to the participants (56%) than the non-preferred ones (22% each). Furthermore, the amount of visual information was high enough (1880 ms) to ensure that participants correctly inferred the different intentions and could therefore be biased towards the intention with the highest probability. The order of presentation of the trials was pseudo-randomized within the block.

2.2: *covert* blocks of movies with equal probabilities.

As for the baseline session, each covert block involved two different amounts of information ($Inf[1,2]$, $Inf[1,3]$, or $Inf[2,3]$) presented in a pseudo-random order. However, in these blocks, the motor intentions had equal probability.

The reason for block interleaving was that it enabled us to maintain the bias constant across the bias session. Indeed, by regularly inserting overt blocks of movies with different probabilities, we insure that participants were continuously biased towards one intention over the whole session. Furthermore, even though the baseline session did not include any bias assignment, and therefore was not concerned with bias maintenance, it contained the same trial organization (block interleaving) to allow a direct comparison of the performance between the two sessions.

Training session

Prior to running the experiment, participants undertook a training session *to get familiar with the task*. The training consisted of 3 baseline experimental sequences (non-biased probability distribution) with interleaved blocks. The 72 movies (3x24) presented during the training session were distinct from those used in the experiment.

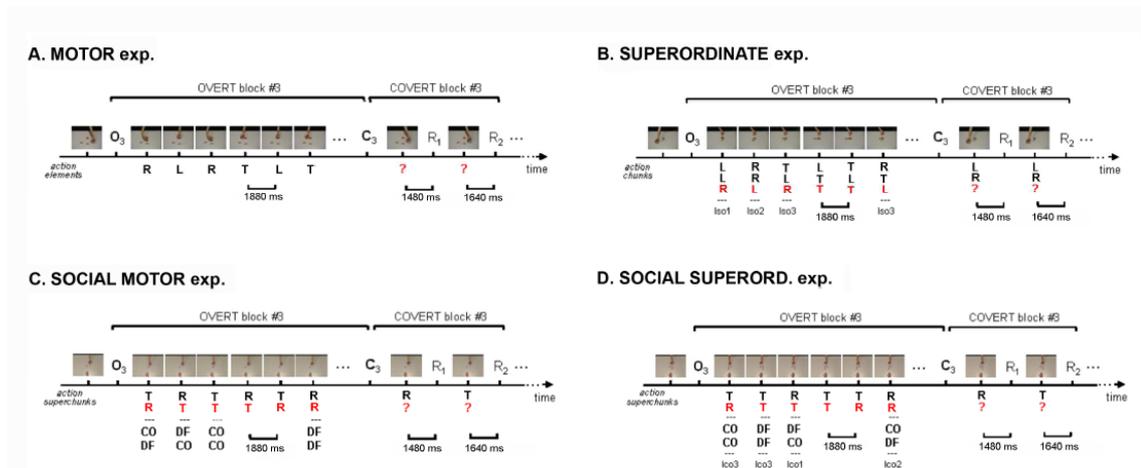


Fig . 2. Experimental designs for the 4 experiments. Examples of a typical experimental sequence (one overt block followed by one covert block) used in both the baseline and the bias sessions. Overt blocks: movies with a very high and constant amount of visual information. Covert blocks: movies with lower and varying amounts of visual information. In the 4 experiments, the probability of all intentions was held equal across the block, except in overt blocks of the bias session, where one particular intention had a greater probability to be accomplished than the other ones. Labels: **L**: “lift” action; **R**: “rotate” action; **T**: “transport” action. **Iso**: shape. **Ico**: configuration. **CO**: cooperate; **DF**: defect.

Design and statistical analyses

One group of statistical analyses was performed for each session independently (baseline and bias sessions), on the two dependant variables (participants’ hits and reaction times for correct responses). In the overt blocks, one two-tailed t-test was conducted, for each session, on participants’ reaction times (RTs) between non-preferred and preferred intentions. In the covert blocks, a 2x3 repeated-measures ANOVA was performed for each session on both

RTs and hits. The first two-level factor was the bias (preferred vs. non-preferred intention) and the second three-level factor was the amount of visual information (Inf1, Inf2 and Inf3). Post-hoc Fisher tests were then performed to identify differences between conditions.

Another group of analyses was conducted in order to assess the magnitude of the bias effect on participants' performance. To do so, we looked at whether increasing the probability of one intention concomitantly affected the selection of intentions with lower probability. Two-tailed t-tests on RTs and hits for non-preferred intentions were thus performed between the baseline and the bias sessions. We predicted that selecting a non-preferred intention should be more demanding in the bias session – as it concomitantly requires inhibiting a competing biased intention – than in the baseline session, where all intentions had the same probability to occur. In the following, the resulting “cost” (i.e. increased RTs and decreased hits for non-preferred intentions) was considered as an indirect measure of the bias effect.

For all the analyses, a $p < .05$ was taken as the criterion for significance and an eta squared (η^2) was used as a measure of effect size. These analyses were performed using the statistical software *Statistica 7* (www.statsoft.com).

2.1.1.2. Results

For each session, two-tailed t-tests were performed between the two non-preferred intentions on both RTs and hits. As no significant differences appeared (all $p > .05$), performances for these two non-preferred intentions were pooled across for subsequent analyses.

Overt blocks

As expected, participants performed well on the task when the amount of visual information was very high (mean correct responses percentage = 98%, S.D. = 2.4, and 96.8%, S.D. = 3.4 in the baseline and the bias sessions, respectively). Furthermore, in the baseline session, there were no significant differences among hits and RTs between the (future) preferred intention (i.e. the one towards which participants will be biased in the bias session) and the non-preferred intention, indicating that prior to biasing participants, there wasn't any *a priori* bias towards one intention rather than another one (two-tailed *t*-tests, all $p > 0.2$, see figure 3, 'Motor exp.', baseline session).

The only significant difference was found in the bias session, with faster RTs for the preferred motor intention vs. non-preferred intentions (two-tailed *t*-tests, all $p < .001$). Subsequent analyses of RTs across time were carried out by independently comparing RTs for preferred motor intention and RTs for non-preferred motor intentions across the different blocks. The bias was found to have a cumulative effect over time, with RTs for preferred intention progressively decreasing up to the block 8 (minimal RT=337ms) and then remaining constant until the end of the session (blocks 1-3 vs. blocks 4-6: $t=3.09$, $p < .005$; blocks 4-6 vs. blocks 7-9: $t=2.08$, $p < .05$) (see figure 3, 'Motor exp.').

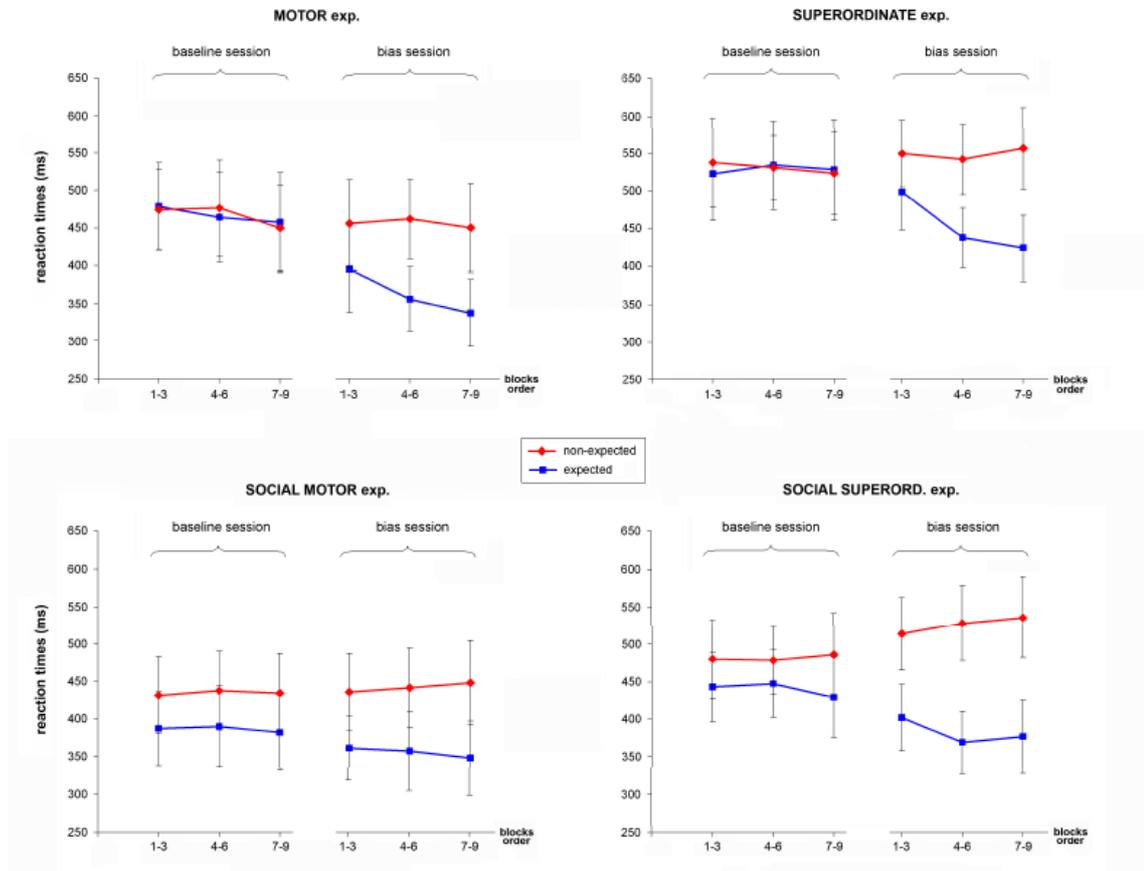


Fig . 3. Mean reaction times (\pm SD) of the non-preferred and the preferred intentions across time in the OVERT blocks for the baseline and the bias sessions. (1-3): the first three overt blocks of the session; (4-6) the intermediate three blocks; and (7-9) the three last blocks of the session.

Covert Blocks

The 3 (amounts of information) x 2 (preferred vs. non-preferred intentions) repeated-measures ANOVA revealed significant effects on both participants' RTs and hits. In the baseline session, a significant effect of the amount of visual information was obtained (RTs: $F(2,116)=167.13$, $p<.001$, $\eta^2 = 0.74$, and hits: $F(2,116)=277.44$, $p<.001$, $\eta^2 = 0.82$). As expected, RTs were found to decrease and hits to improve as the amount of visual information increased. There were, however, no significant effect of the bias nor of the interaction between the amount of visual information and the bias (both $p>0.05$), showing that improved participants' performance for higher amounts of information was independent of the presented intention (Im1, Im2, or Im3) (see table 1 and figure 4, 'Baseline session').

In the bias session, in addition of an effect of the amount of visual information (RTs: $F(2,116)=98.8$, $p<.001$, $\eta^2 = 0.62$; hits: ($F(2,116)=190.92$, $p<.001$, $\eta^2 = 0.76$), an effect of the bias (RTs: $F(1,58)=18.51$, $p<.001$, $\eta^2 = 0.24$; hits: ($F(1,58)=19.81$, $p<.001$, $\eta^2 = 0.25$) as well as of the interaction (RTs: $F(2,116)=13.98$, $p<.001$, $\eta^2 = 0.19$; hits: $F(2,116)=5.44$, $p=.005$, $\eta^2 = 0.08$) were also observed. Post-hoc tests indicated that participants were more accurate and faster in recognizing the preferred intention in condition of low information only (LSD Fisher tests, RTs: $\text{Inf1} = p<.001$; $\text{Inf2} = p<.005$ and $\text{Inf3} = ns$; Hits: $\text{Inf1} = p<.001$; $\text{Inf2} = ns$. and $\text{Inf3} = ns$.) (see table 1 and figure 4, 'Bias session').

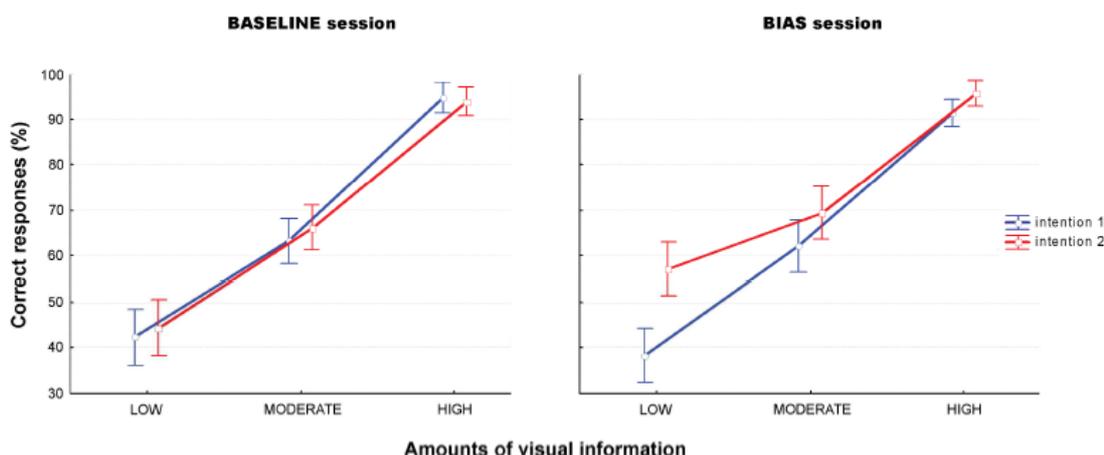


Fig . 4. Non-social motor intention experiment (COVERT blocks): mean percentage of correct responses (\pm SD) for preferred (red) and non-preferred (blue) intentions for each amount of visual information (low, moderate, high).

Table 1. Non-social motor intention experiment (COVERT blocks): mean reaction times (\pm SD) for preferred and non-preferred intentions for each amount of visual information (low, moderate, high). P: preferred intention; n-P: non-preferred intention.

Experiment Reaction Times (ms)	
MOTOR	LOW MODERATE HIGH
Session	
Baseline intention 1	1153 \pm 303 894 \pm 205 579 \pm 193
intention 2	1092 \pm 288 866 \pm 247 601 \pm 143
Bias intention 1 (n-P)	1190 \pm 380 975 \pm 294 628 \pm 206
intention 2 (P)	817 \pm 270 659 \pm 223 552 \pm 159

Effect of the bias on the non-preferred intention

We were also interested in evaluating the influence of the bias on the selection of the non-preferred intentions. Comparing the performance for the non-preferred intention between the two sessions revealed no significant differences for any amount of visual information (two-tailed t-tests, RTs and hits: all $p>0.05$). This result indicates that switching from the baseline to the bias session (i.e. increasing the probability of one intention to the detriment of others) did not significantly affect the inference of motor intentions with lower probabilities.

2.1.1.3. Preliminary discussion

In the present experiment, participants were instructed to infer the *motor intention* (to lift, rotate or transport) of an actor manipulating a meaningless object. The participants' inference was constrained by two factors that varied over the blocks: the amount of visual information available from the action scene and the prior probability for a given intention to be accomplished. Three main results emerged. First, as expected, motor intentions were better inferred when the actions were presented with a high amount of visual information. Second, performances were also influenced by the probability distribution of the intentions, with a significant decrease of participants' error rates and RTs towards the preferred (i.e. biased) motor intention. This bias effect may be accounted for by the existence, in participants, of a privileged representation of the most frequent intention. Held active all over the block and refreshed from block to block, this representation would bias the selection of one intention by continuously exerting top-down influences on participants' choice. We also found that the bias significantly interacted with the amount of visual information. Indeed, when participants were shown an action scene with a low amount of visual information, making the intention hardly recognizable, their responses preferentially shifted towards the bias (i.e. they responded more towards the preferred intention). In this condition, participants were mostly relying on their prior expectations to infer the intention underlying the action scene. Crucially, however, this shift in the participants' response was only observed in the condition of low amount of visual information. Indeed, as this amount increased, the bias had less effect on the participant's performance. This result may reflect the specificities of motor intentions. In the motor condition, the visual information directly denotes the content of the underlying intention. When participants are exposed to a sufficient amount of visual information, the prior expectations tend to exert increasingly less influence on perception because, in this case, the visual information is reliable enough for correctly inferring the actor's current intention.

2.1.2. Experiment 2: non-social superordinate intention

2.1.2.1. Methods

Participants

30 new participants (15 males, 15 females, mean age = 36.59, S.D. = 8.12 and laterality score mean = 0.79, S.D. = 0.19; Oldfield, 1971) took part in this experiment. They all reported normal or corrected-to-normal visual acuity and received 10 euros for taking part in the study.

Stimuli

As with the non-social motor experiment, test material consisted of incomplete movies clips showing an actor's hand manipulating meaningless objects. However, contrary to experiment 1 in which one single motor act was shown to the participants, movies in the superordinate experiment represented a sequence of three successive manipulations (to transport, rotate or lift the cube) leading to the construction of a meaningless shape. Each sequence was therefore characterized by an underlying superordinate intention, represented by one final shape (Iso1: shape S1, Iso2 or Iso3). The objects of the first experiment were also used in this second experiment (figure 1, B). The first action was performed on one of the three cubes, the second action on one of the two cubes left, and the third action on the last cube left. After each action, the hand came back to the starting position. The incompleteness of the video sequences was controlled so that the duration of

the last action was varied on 3 distinct levels (*i.e.* 1480 ms, 1560 ms or 1640 ms after this action starts). All the movies were realized with the same actor as for experiment 1.

Procedure

The organization of the trials was of the same as in the non-social motor experiment (see figure 2, B, 'Superordinate exp.'). For each trial, participants were requested to observe an incomplete movie and infer what the final (*i.e.* superordinate intention) was. However, to prevent participants from adopting alternative strategies, the task was further constrained. First, to ensure that participants paid attention to the overall sequence of actions and did not just passively recognize the final shape, they were asked to make their response on the last action of the sequence. They indicated what the not-yet completed action of the sequence was, by pressing as quickly and as accurately as possible, one of the three corresponding keyboard buttons (T, L, or R). This response, therefore, directly depended upon having correctly inferred the superordinate intention of this sequence. For example, responding 'lift' after two successive 'rotate' actions required having inferred that the superordinate intention was the shape *iso1* and that the construction of this shape necessitated that the last action be a 'lift' one. Second, to ensure that participants were biased towards the superordinate intention itself and not towards the last final action only, commutative sequences were used so that each shape could be constructed from distinct sequences of actions. Sequences shown in the covert blocks were thus distinct from those used in the overt blocks (*e.g.* the shape S1 could be obtained from the sequence 'lift-lift-rotate' in an overt block, but from the sequence 'lift-rotate-lift' in a covert block). Finally, the intention could not be inferred solely from the motor acts composing the sequence. Indeed, although the probability for each shape to be constructed was manipulated in the bias session (*i.e.* one particular shape had a higher probability), the probabilities of the individual actions were held equal at each step (*e.g.* steps 1, 2 and 3) of that sequence.

As in experiment 1, all the trials were equalized for temporal homogeneity (see Appendix) and each movie was presented once.

Training session

Participants undertook a training session prior to running the experiment. *The training consisted of 3 baseline sequences (non-biased probability distribution) with interleaved blocks.* The 72 movies (3x24) presented during the training session were distinct from those of the experiment.

2.1.2.2. Results

Statistical analyses were similar to those conducted in the first experiment. Responses for the two non-preferred superordinate intentions were pooled across for subsequent analyses as there were no significant differences among both hits and RTs between these responses (for each session, two-tailed *t*-tests, $p > 0.15$).

Overt blocks

As for experiment 1, participants performed well on the task when the amount of visual information was very high (1880 ms), in both the baseline (mean correct responses percentage: 98.6%, S.D. = 2.3) and the bias session (mean correct responses percentage: 98.1%, S.D. = 3.1). In the baseline session, furthermore, participants were equally rapid

at inferring the last action of the sequence, whatever the superordinate intention being accomplished (Iso1, Iso2, or Iso3) (two-tailed t -tests, all $p > .35$).

In the bias session, however, although there were no significant differences among hits between preferred and non-preferred intentions, RTs for the preferred intention were found to significantly decrease (two-tailed t -tests, preferred vs. non-preferred intentions, all $p < .001$). This decrease accentuated over time as revealed by a cumulative effect of the bias across blocks. Indeed, RTs decreased up to the block 7 (minimal RT=424ms) and then remained constant until the end of the session (blocks 1-3 vs. blocks 4-6: $t=4.04$, $p < .001$; blocks 4-6 vs. blocks 7-9: $t=0.83$, $p > .05$) (see figure 3, 'Superordinate exp.').

Covert blocks

In both the baseline and the bias sessions, the amount of visual information significantly affected participants' hits and latencies with decreased RTs (baseline: $F(2,116)=423.68$, $p < .001$, $\eta^2 = 0.87$; bias session: ($F(2,116)=523.9$, $p < .001$; $\eta^2 = 0.9$) and a greater number of hits as the amount of visual information increased (baseline: $F(2,116)=249.18$, $p < .001$, $\eta^2 = 0.81$; bias session: ($F(2,116)=199.03$, $p < .001$; $\eta^2 = 0.77$) (see table 2 and figure 5).

In the bias session, a significant effect of the bias was also observed with faster RTs and increased hits for actions congruent with the preferred intention (RTs: $F(1,58)=47.04$, $p < .001$, $\eta^2 = 0.44$; and hits: $F(1,58)=62.09$, $p < .001$, $\eta^2 = 0.51$). Finally, the bias was found to significantly interact with the amount of visual information in such a way that the number of hits was significantly higher and RTs faster for the preferred intention as the amount of visual information decreased (RTs: $F(2,116)=15.3$, $p < .001$, $\eta^2 = 0.2$; and hits: $F(2,116)=9.28$, $p < .001$; $\eta^2 = 0.13$). Post-hoc tests further indicated that participants were more accurate and faster in recognizing the preferred intention in condition of both low and moderate amounts of visual information (LSD Fisher: $Inf1$, $p < .001$; $Inf2$, $p = .005$, $Inf3$, $p < .05$ for RTs; and $Inf1$, $p < .001$; $Inf2$, $p < .001$; $Inf3 = ns.$ for hits) (see table 2 and figure 5, 'Bias session').

Effect of the bias on the non-preferred intention

The number of correct responses for non-preferred intention significantly decreased in the bias session, compared to the baseline session, for both low and moderate amounts of visual information (two-tailed t -tests: all $t(30) > 2.33$, all $p < 0.02$). Likewise, RTs significantly decreased in the condition of moderate amount of visual information (two-tailed t -tests: $t(30) = -2.09$, $p = 0.04$).

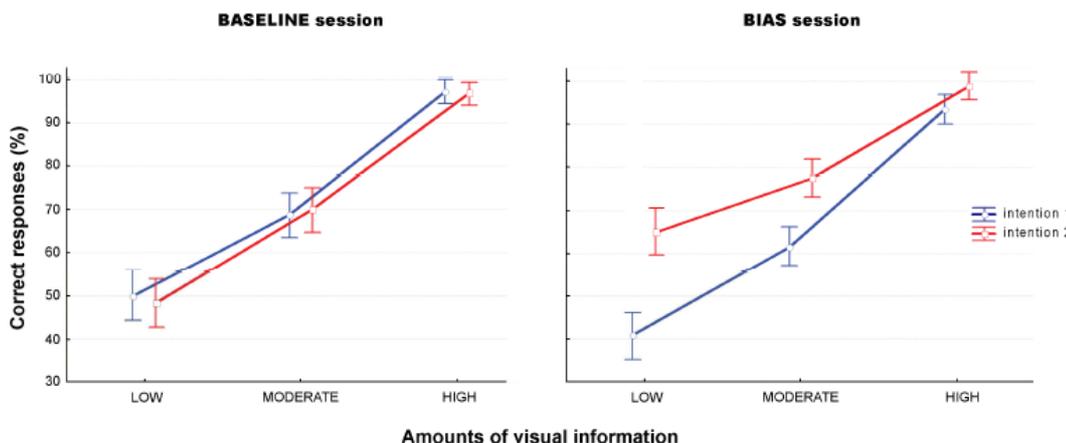


Fig. 5. Non-social superordinate intention experiment (COVERT blocks): mean percentage of correct responses (\pm SD) for preferred (red) and non-preferred (blue) intentions for each amount of visual information (low, moderate, high).

Table 2. Non-social superordinate intention experiment (COVERT blocks): mean reaction times (\pm SD) for preferred and non-preferred intentions for each amount of visual information (low, moderate, high). P: preferred intention; n-P: non-preferred intention.

Experiment Reaction Times (ms)			
	SUPERORDINATE	LOW	MODERATE HIGH
Session			
Baseline intention 1	1605 \pm 314	1241 \pm 218	745 \pm 153
intention 2	1676 \pm 304	1211 \pm 244	783 \pm 138
Bias intention 1 (n-P)	1689 \pm 307	1357 \pm 262	809 \pm 161
intention 2 (P)	1221 \pm 209	1014 \pm 183	600 \pm 138

2.1.2.3. Preliminary discussion

In this second experiment, participants were instructed to infer the non-social superordinate intention (i.e. the final shape) underlying a sequence of actions and to give a response about the last action of that sequence. As in the previous experiment, participants' error rates and latencies decreased as the amount of visual information increased. Furthermore, the bias was also found to significantly affect performance, with participants being more accurate and faster at recognizing actions achieving a preferred superordinate intention. This bias effect was accounted for by a preference towards the superordinate intention itself rather than towards a single motor act of the sequence since only the probability of one superordinate intention was increased, whereas the probability of each single action was maintained equal over the blocks. The bias effect was also observed to significantly interact with the amount of visual information. The less visual information participants had about the action scene, the more they relied on their priors, with their responses exhibiting a massive shift towards actions achieving a preferred superordinate intention.

However, two additional results make the present experiment diverge from the previous one. First, the bias effect was greater in the second experiment, as it was observed both in the conditions with low and moderate amounts of visual information, whereas, in the motor task, the bias affected participants' performance only in the condition of low visual information. This over-reliance on prior expectations is likely to be accounted for by the very properties of the superordinate intention. The achievement of a superordinate intention indeed requires the execution of a sequence of motor acts, each of which being guided by a corresponding motor intention. As superordinate intentions can be accomplished by *different* combinations of motor acts, the adequate strategy for participants may thus consist in maximising their prior expectations— i.e. in privileging responses congruent with their prior expectations – even in conditions where the amount of visual information was moderate.

Importantly, this observation is further supported by the 'switch effect' observed in conditions of low and moderate amounts of visual information. Indeed, together with an increasing number of correct responses and decreasing latencies for the preferred intention, decreased hits and increased RTs were recorded for the non-preferred intentions. This effect reflects the cost associated with a switch from a privileged (preferred intention) to a non-privileged representation (non-preferred intention), and as such, may be considered as a consistent measure of the bias effect. When inferring a superordinate intention, the greater extent to which participants rely on the bias thus resulted in an increasing difficulty in

disengaging from their prior expectations to select a less privileged representation of the intention.

2.2. Social experiments

The third (fig. 1, **C**) and fourth experiments (fig. 1, **D**) presented two actors engaged in a social game in which they could either cooperate or defect. Participants were instructed to infer the nature of the second player's social intention (i.e. cooperative or defective intention). In both these experiments, the bias was assigned according to the way the second player responded to the strategy adopted by the opponent in the previous round. Participants were therefore biased towards the reputation of the second player rather than towards one particular type of social intention. Finally, as in the two previous non-social experiments, both *motor intentions* (single motor acts) and *superordinate intentions* (sequences of motor acts leading to the construction of a shape) were considered in the last two experiments.

2.2.1. Experiment 3: social motor intention

2.2.1.1. Methods

Participants

30 novel participants (15 males, 15 females, mean age = 32.9, S.D. = 10, and laterality score mean = 0.80, S.D. = 0.11; Oldfield, 1971) participated in this experiment. They all reported normal or corrected-to-normal visual acuity and received 10 euros for taking part in the study.

Stimuli

Visual stimuli consisted of incomplete movies showing two players' hands (one in front of the other) manipulating meaningless objects. Two cubes (printed with a blue or a red line) were disposed on the sides of a grid that was situated in the centre of the scene (figure 1, C). The cubes were of similar size (6cm X 6cm) and located at equal distance from the starting position of each player's hand (figure 1, C). The two actors played one after the other by moving the cube towards the middle of the grid (termed 'bank') or by rotating it so that it remains at its own place. Movies were cut so that the last action (i.e. the action performed by the second player) was incomplete (Inf1=1480 ms, Inf2=1560 ms, Inf3=1640 ms after the last action starts).

Here, each motor act directly denoted the social intention of the player: each player could either cooperate with the other one, by moving the cube towards the central bank (transport), or defect, by leaving the cube at its own place (rotate). Consequently, there were four possible combinations of intentions, or strategies: either both players cooperated (transport/transport) or defected (rotate/rotate), or the first player defected and the second cooperated (rotate/transport), or the first player cooperated and the second defected (transport/rotate).

Procedure

For each trial, participants were instructed to observe the incomplete movie and infer what the last action (i.e. the one performed by the second player) was. This response required to have inferred the second player's intention (to defect or to cooperate) which itself depended upon the first player's strategy. Participants were asked to give their answer by pressing, as quickly and as accurately as possible, one of the two keyboard buttons corresponding

to the two possible last actions (T for 'transport', or R for 'rotate') susceptible to achieve the second player's intention. Once a response was given, the next trial started.

In the baseline session, all combinations of strategies were counterbalanced over the blocks (i.e. whatever the first player did the second player could as much defect as cooperate). In the bias session, on the other hand, the probability that the second player does whatever the opponent has done in the previous round was increased, thus biasing participants to perceive the second actor as a "tit-for-tat" player (i.e. as being more inclined to cooperate if the first player had previously cooperated, and to defect if the first player had previously defected). The rationale for biasing the second player's reputation in such a way is threefold. First, tit-for-tat reputation (TFT) implies that individuals respond to their opponent's actions in a mirrored (i.e. correlated) fashion. Therefore, successfully predicting intentions of a TFT-like player necessarily involves taking into account what the first player has done, and by consequence, ensured that participants paid attention to the whole sequence of actions (both actor 1's play and actor 2's play). Second, contrary to other common types of reputation such as "always defect", or "always cooperate", TFT may equally imply cooperative as well as defective strategies. The probability that the second actor behaves as a TFT player could thus be increased without otherwise increasing the probability of one intention (e.g. cooperate) to the detriment of the other one (e.g. defect). Holding equal the probability of both these strategies was here crucial to nullify their potential kinematics differences on participants' performance (see Georgiou, Becchio, Glover, & Castiello, 2007) and to avoid stereotyped responses as well (e.g. always responding 'cooperate' or 'defect'). Finally, in free-error environment, a tit-for-tat strategy is known to be a more intuitive and successful strategy than alternative ones, such as "always cooperating", "always defecting" or "acting randomly" (Rapoport & Chammah, 1965; Axelrod, 1997; André & Day, 2007). We thus chose to experimentally strengthen this already existing *a priori* bias by increasing, over the blocks, the probability that the second player's action mirrors her opponent's one, while keeping equal both the probability of each single act (to rotate or to transport) and the overall probability of each intention (to defect or to cooperate) across the blocks (figure 2, C, 'Social Motor exp.').

All the trials were equalized for temporal homogeneity (see Appendix) and each movie was presented once across the whole experiment.

Training session

As in the previous experiments, participants undertook a training session (*three baseline sequences with interleaved blocks*) prior to running the experiment. The 72 movies presented during the training session were distinct from those of the experiment.

2.2.1.2. Results

Statistical analyses were similar to those conducted in the previous experiments. Two-tailed *t*-tests revealed no significant differences among both participants' RTs and hits between the TFT intentions (coop/coop vs. def/def: two-tailed *t*-tests, all $p > .2$) and between the alternative ones (coop/def vs. def/coop, two-tailed *t*-tests, all $p > .15$). Performances for def/def were therefore pooled with those of coop/coop (i.e. TFT or preferred intentions) and performances for def/coop were pooled across with that of coop/def (i.e. alternative or non-preferred intentions) for the subsequent analyses.

Overt blocks

Participants performed well on the task in both the baseline and the bias sessions (mean correct responses percentage: 98%, and S.D. = 2.5 and S.D. = 2.8). In the baseline session, RTs analyses revealed a significant effect of the type of reputation, with participants being faster at inferring an action that was embedded within a tit-for-tat strategy than within an alternative strategy (e.g. always defect or always cooperate) (two-tailed *t*-tests, all $p < .05$). This result confirmed the existence of an inherent preference towards TFT reputation. This effect was maintained after the bias assignment as revealed by faster RTs for the preferred intention (i.e. the TFT intention) in the bias session (two-tailed *t*-tests, preferred vs. non-preferred intentions, all $p < .001$). However, RTs for preferred intentions did not significantly decrease with time (blocks 1-3 vs. blocks 4-6: $t = 0.35$, $p > .05$; blocks 4-6 vs. blocks 7-9: $t = -0.68$, $p > .05$) (see figure 3, 'Social Motor exp.').

Covert blocks

Baseline session

ANOVAs performed on both hits and RTs showed a significant main effect of the amount of visual information with decreased RTs ($F(2,116) = 80.44$, $p < .001$; $\eta^2 = .58$) and a greater number of hits ($F(2,116) = 209.02$, $p < .001$, $\eta^2 = .78$) as the amount of visual information increased. In addition, the main effect of the type of reputation – previously observed in the overt blocks – was also significant in the covert blocks among RTs only ($F(1, 58) = 4.7$, $p < .05$, $\eta^2 = .07$). The second player's action was more rapidly inferred when it was embedded within a TFT intention than within an "always defecting" or an "always cooperating" intention. The type of reputation did not, however, interact with the amount of visual information (see table 3 and figure 6, 'Baseline session').

Bias session

There were significant main effects of the amount of visual information and of the bias on both the RTs (main effect of amount of visual information: $F(2,116) = 114.49$, $p < .001$, $\eta^2 = .66$; main effect of the bias, $F(1,58) = 25.29$, $p < .001$, $\eta^2 = .3$) and the hits (main effect of amount of visual information: $F(2,116) = 170.34$, $p < .001$, $\eta^2 = .74$; main effect of the bias, $F(1,58) = 34.75$, $p < .001$, $\eta^2 = .37$), as well as a significant effect of the interaction between these two factors (RTs: $F(2,116) = 9.23$, $p < .001$, $\eta^2 = .13$; hits: $F(2,116) = 8.28$, $p < .001$, $\eta^2 = .12$). Participants' performance (slower RTs and higher hits) for actions congruent with a preferred social intention improved as the amount of visual information decreased. Furthermore, the bias significantly affected participants' hits for all amounts of information (LSD Fisher: $\text{Inf}1$, $p < .001$; $\text{Inf}2$, $p = .005$; $\text{Inf}3$, $p = ns$ for RTs; $\text{Inf}1$, $p < .001$; $\text{Inf}2$, $p = .05$; $\text{Inf}3$, $p = .05$ for hits) (see table 3 and figure 6, 'Bias session').

Effect of the bias on the non-preferred intention

When comparing performance for non-preferred intentions between the baseline and the bias sessions, we found significant differences between these sessions for a high amount of visual information only, with participants' RTs for non-preferred intentions significantly decreasing in this condition (two-tailed *t*-tests: $t(30) = 2.26$, $p = 0.03$).

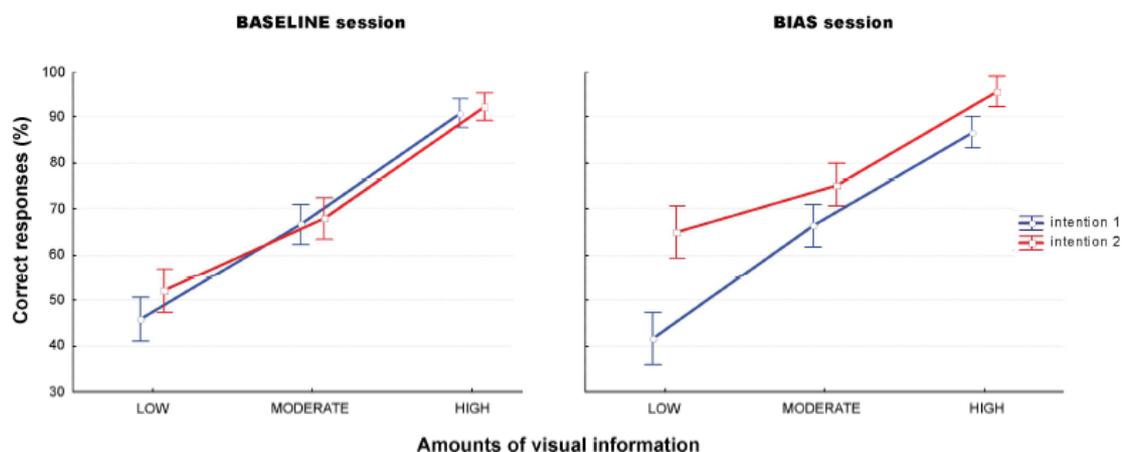


Fig. 6. Social motor intention experiment (COVERT blocks): mean percentage of correct responses (\pm SD) for preferred (red) and non-preferred (blue) intentions for each amount of visual information (low, moderate, high).

Table 3. Social motor intention experiment (COVERT blocks): mean reaction times (\pm SD) for preferred and non-preferred intentions for each amount of visual information (low, moderate, high). P: preferred intention; n-P: non-preferred intention.

Experiment Reaction Times (ms)	
SOCIAL MOTOR	LOW MODERATE HIGH
Session	
Baseline intention 1	1159 \pm 263 990 \pm 220 888 \pm 210
intention 2	1053 \pm 232 899 \pm 213 750 \pm 169
Bias intention 1 (n-P)	1217 \pm 297 1041 \pm 282 788 \pm 229
intention 2 (P)	846 \pm 220 748 \pm 216 607 \pm 158

2.2.1.3. Preliminary discussion

The motor social experiment presented two actors engaged in a social game. Participants were instructed to infer the action (to rotate or to transport) accomplishing the social intention (to cooperate or to defect) of the second player, given the type of strategy previously adopted by the first player. As in the two previous experiments, the participants' performance was constrained by two factors that varied over the blocks: the amount of visual information available from the scene and the probability that one particular social intention was the real cause of the observed action.

Most of the results were similar to those obtained in the two previous experiments. Participants were better at inferring the action as the amount of visual information increased. Likewise, performances were affected by the probability distribution of the sequences, with a significant preference for actions associated with the intention towards which they were previously biased (i.e. TFT intention). Indeed, participants were more inclined to infer that the second player would defect (or cooperate) if the first player had previously defected (or cooperated). It is noteworthy that the bias observed in participants' responses can only be accounted for by a bias towards TFT intentions themselves, rather than towards one specific motor act or one particular type of social strategy (defective or cooperative), the probability of which did not vary across blocks (i.e., each player performed the action 'to rotate' as

often as the action 'to transport'; likewise, each actor cooperated as often as he defected). This observation indicates that, as expected, participants' inference of the current action was primarily driven by the nature of the social interaction between players, rather than by a preference for one specific motor act or one single strategy. Finally, this preference for TFT intentions interacted with the amount of visual information in such a way that participants relied more on their prior expectations as the amount of visual information decreased.

Two additional results were, however, found, that did not appear in previous experiments. First, in the baseline session, where no bias was assigned, participants were yet faster at predicting an action associated with a TFT intention. Interestingly, this early preference for TFT strategies differed from several aspects of the probabilistic bias that was imposed on participants in the second session. Not only this preference did not interact with the amount of visual information but its effect on performance remained constant over time. Second, the effect of the probabilistic bias was significant for any amount of visual information (low, moderate and high), showing that the prior expectations contributed to the inference even in conditions in which the visual information was highly reliable.

Reputational effects may, at least in part, account for these two last results (Frith & Firth, 2006). Knowledge about an individual's reputation indeed provide strong expectations on how this individual will subsequently behave – as long as she deserves her reputation (Singer et al., 2006; Fu & Lee, 2007). In the present task, observing repeated interactions between both actors participated in creating such prior expectations, then biasing participants' prediction about which kind of strategy (to cooperate or to defect) the second player would be more likely to adopt, given the reputation she progressively acquired across the blocks. The bias effect precisely accounts for the fact that participants' inference did integrate this reputational knowledge. By increasing the probability that the second player does whatever the opponent has done in the previous round, one forced participants to perceive the second player as a 'tit-for-tat' player, rather than an altruist (always cooperate), an egoistic (always defect), or a 'random' player, hence generating progressively reinforced expectations that may come to prevail on relevant perceptual cues – as it is the case in conditions of high amount of visual information.

2.2.2. Experiment 4: social superordinate intention

2.2.2.1. Methods

Participants

30 novel participants (15 males, 15 females, mean age = 34.27, S.D. = 9.42, and laterality score mean = 0.83, S.D. = 0.26; Oldfield, 1971) participated in this last experiment. They all reported normal or corrected-to-normal visual acuity and received 10 euros for taking part in the study.

Stimuli

Participants were observing incomplete movies clips showing two players' hands (one in front of the other) manipulating meaningless objects (cubes printed with a blue or a red line). Four cubes (one blue and three red ones) were placed in the centre of the scene. Actors played in turn with the goal to vertically align three cubes among the four available ones (see figure 1, D). The goal of the first player was to align the three cubes according to the color red, irrespective of the orientation (configuration 1 or *lco1*). The goal of the second player was to align the cubes according to the orientation, irrespective of the color (configuration 2

or *Ico2*). Finally, a third configuration could be obtained by aligning the cubes according to both the orientation and the color (configuration 3 or *Ico3*).

The cubes were of similar size (6cm X 6cm) and placed at equal distance from the starting position of each player's hand (figure 1, D). The first player manipulated the cube closest to him, and then the second player in turn manipulated the cube closest to him. To win the game, each player depended upon the other player's manipulation. As in the social motor experiment, the two social intentions were of a defective or a cooperative nature. However, in the present experiment, the social intention was denoted by the sequence of the players' motor acts (i.e. the final configuration: *Ico1*, 2 or 3), rather than by the single action performed by each player. Each configuration was associated with a defective or a cooperative strategy that could be equally achieved by rotating or transporting the cube. Each player could therefore adopt a defective strategy by manipulating the cube in such a way that it prevented the creation of a configuration, or adopt a cooperative strategy in order to achieve a configuration. The resulting four possible combinations of strategies therefore led to different final configurations: either both players defected or cooperated (*Ico3*), or the first player defected and the second cooperated (*Ico1*), or the first player cooperated and the second defected (*Ico2*). As for previous experiments, only the last action (i.e. the action performed by the second player) was made incomplete ($Inf1=1480$ ms, $Inf2=1560$ ms, $Inf3=1640$ ms).

Procedure

Participants were instructed to infer the social superordinate intention of the second player by indicating which action allows the accomplishment of that intention. A correct response thus required having correctly inferred the second player's intention (to defect or to cooperate) which itself depended upon what the first player had previously done. Participants were asked to give their answer by pressing, as quickly and as accurately as possible, one of the two keyboard buttons corresponding to the two possible actions (T for 'transport', or R for 'rotate') congruent with the second player's social intention (def. or coop.). Once a response was given, the next trial started.

The organization of trials in the social superordinate experiment was the same as for the social motor experiment (see figure 2, D). Tit-for-tat reputation (i.e. the second player did whatever the opponent has played in the previous round) was chosen to be biased across participants. Likewise, commutative sequences were also used so that each pattern could be obtained from distinct sequences of actions (e.g. 'transport' and 'rotate' actions could equally achieve an *Ico1*, *Ico2* or *Ico3* sequence). This ensured that the different strategies could not be predicted from motor acts solely. Finally, both the overall probability of each type of strategy (cooperative or defective) and each individual action (to rotate, to transport) were keeping equal across the blocks, while increasing the probability that the second player's action mirrors her opponent's one (coop/coop, def/def > coop/def, def/coop).

All the trials were equated for temporal homogeneity (see Appendix) and each movie was presented once.

Training session

As in experiment 1, participants undertook a training session prior to running the experiment. *The training consisted of 3 baseline sequences with interleaved blocks.* The 72 movies presented during the training session were distinct from those used in the experiment.

2.2.2.2. Results

Statistical analyses were similar to those conducted in the first three experiments. Two-tailed *t*-tests revealed no significant differences among both participants' RTs and hits between the two preferred combinations (coop/coop vs. def/def: two-tailed *t*-tests, all $p > .2$) and between the two non-preferred combinations (coop/def vs. def/coop, two-tailed *t*-tests, all $p > .15$). Performances for def/def were therefore pooled across with that of coop/coop (TFT or preferred intentions) and performances for def/coop were pooled across with that of coop/def (non-preferred intentions).

Overt blocks

Participants performed well on the task in both the baseline (mean correct responses percentage: 97.5%, S.D. = 2.8) and the bias sessions (mean correct responses percentage: 98.4%, S.D. = 2.6). In the baseline session, RTs analyses revealed a significant effect of the type of reputation (two-tailed *t*-tests, all $p < .05$), with participants being faster at inferring an action that was embedded within a tit-for-tat strategy than within an alternative strategy (i.e. always defect or always coop). RTs for the preferred intention also significantly decreased in the bias session (two-tailed *t*-tests, preferred vs. non-preferred intentions, all $p < .001$) and this decrease was found to accentuate over time up to the block 6 (minimal RT=369 ms), (blocks 1-3 vs. blocks 4-6: $t = -2.11$, $p < .05$; blocks 4-6 vs. blocks 7-9: $t = -0.44$, $p > .05$) (see figure 3, 'Social superord. exp.').

Covert blocks

Baseline session

As the amount of visual information increased, decreased RTs ($F(2,116) = 93.23$, $p < .001$, $\eta^2 = .61$) and increased hits ($F(2,116) = 281.6$, $p < .001$, $\eta^2 = .82$) were observed for actions accomplishing the preferred social intention. In addition, as in the overt blocks, actions were better and more rapidly inferred when they were embedded within a tit-for-tat strategy than within an alternative strategy (RTs: $F(1,58) = 4.16$, $p < .05$, $\eta^2 = .06$, and hits: $F(1,58) = 7.96$, $p < .01$, $\eta^2 = .12$). The effect of reputation did not, however, interact with the amount of visual information showing that the type of strategy affected participants' performance independently of the amount of visual information (see table 4 and figure 7, 'Baseline session').

Bias session

There were significant main effects of the amount of visual information and of the bias on both the RTs (main effect of amount of visual information: $F(2,116) = 163.11$, $p < .001$, $\eta^2 = .73$; main effect of the bias $F(1,58) = 52.03$, $p < .001$, $\eta^2 = .47$) and the hits (main effect of amount of visual information: $F(2,116) = 198.77$, $p < .001$, $\eta^2 = .77$; main effect of the bias $F(1,58) = 92.16$, $p < .001$, $\eta^2 = .61$), as well as a significant effect of the interaction between these two factors (RTs: $F(2,116) = 27.74$, $p < .001$, $\eta^2 = .32$, and hits: $F(2,116) = 4.41$, $p = .01$, $\eta^2 = .07$). Participants' performance (slower RTs and higher percentage of hits) for preferred social intention improved to a large extent as the amount of visual information decreased, although the bias effect was observed for all amounts of visual information as revealed by the Post-hoc (LSD Fisher: $\text{Inf}1$, $p < .001$; $\text{Inf}2$, $p < .001$; $\text{Inf}3$, $p = .01$ for RTs and $\text{Inf}1$, $p < .001$; $\text{Inf}2$, $p < .001$; $\text{Inf}3$, $p < .001$ for hits) (see table 4 and figure 7, 'Bias session').

Effect of the bias on the non-preferred intention

We found the reaction times for non-preferred intentions to significantly increase in the bias session, compared to the baseline session, for all amounts of visual information (two-tailed t-tests: all $t(30) < -2.07$, all $p < 0.05$). The number of correct responses for non-preferred intention also significantly decreased in the bias session, for both low and high amounts of visual information (two-tailed t-tests: all $t(30) > 2.33$, all $p < 0.02$). This effect did not reach significance in the condition of moderate amount of visual information (two-tailed t-test: $t(30) = 1.8$, $p = 0.07$).

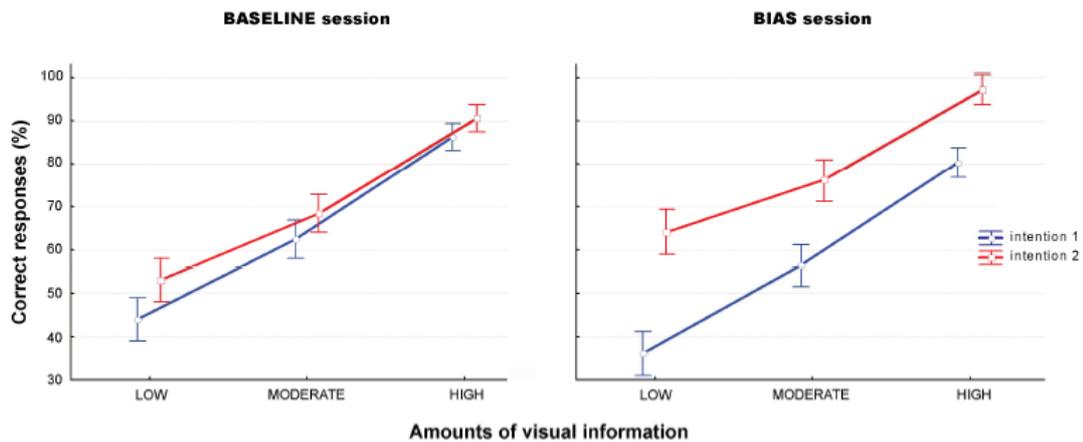


Fig . 7. Social superordinate intention experiment (COVERT blocks): mean percentage of correct responses (\pm SD) for preferred (red) and non-preferred (blue) intentions for each amount of visual information (low, moderate, high).

Table 4. Social superordinate intention experiment (COVERT blocks): mean reaction times (\pm SD) for preferred and non-preferred intentions for each amount of visual information (low, moderate, high). P: preferred intention; n-P: non-preferred intention.

Experiment Reaction Times (ms)	
SOCIAL SUPERORD. LOW MODERATE HIGH	
Session	
Baseline intention 1	1282 \pm 288 1049 \pm 237 877 \pm 198
intention 2	1102 \pm 277 957 \pm 217 750 \pm 175
Bias intention 1 (n-P)	1488 \pm 328 1197 \pm 271 961 \pm 230
intention 2 (P)	923 \pm 216 785 \pm 192 704 \pm 151

2.2.2.3. Preliminary discussion

As in the previous experiment, the current study presented two actors engaged in a social game. The goal of each player was to build a colour or a shape pattern (superordinate intention: Ico1, Ico2, or Ico3) by moving printed cubes on a central grid. Participants were instructed to infer the social superordinate action of the second player (that directly depended upon the strategy adopted by the first player) and give a response about the action accomplishing that intention. The main results paralleled those obtained in the three previous experiments. We found both the amount of visual information and the bias to have a significant effect on participant's performance. As in the third experiment, participants were more accurate and faster at recognizing actions congruent with TFT intentions. Likewise, these performances improved as the amount of visual information increased. It is noteworthy

that, unlike the motor social experiment, the social superordinate intention was specified by the color or the shape pattern (lco1, lco2 or lco3), rather than by one single motor act (to rotate or transport). Indeed, both the probability that one particular action characterizes a cooperative or a defective strategy, and the probability of each strategy to be adopted by one or the other player were counterbalanced across blocks. As a result, a bias observed in participants' responses could only be accounted for by a bias towards a TFT intention (i.e. coop/coop or def/def) since only its probability varied across the blocks. Finally, this preference for TFT intentions was found to interact in the bias session with the amount of visual information, with participants relying more on their prior expectations as this amount decreased.

As for the motor social experiment, the bias effect was of major importance. First, in the baseline session, analyses of latencies and accuracy data revealed that, prior to the bias assignment, participants exhibited an early preference for social intentions congruent with a TFT reputation. Second, in the bias session, the bias significantly impacted on performances for all amounts of visual information. These two results are consistent with the reputational effects that were already observed in the motor social experiment. In the present study, such effects may be accounted for by the fact that the first player's strategy pre-activates, if not reinforces, the expectations of the participant prior observing the second player acting. Here, such reinforcement resulted in an increased difficulty for the observer to disengage from his prior expectations, even in the case where they interfered with the visual cues (e.g. in the conditions of high amount of visual information). Consistent with this interpretation is the observation that participants' hits for the non-preferred intentions were found to significantly decrease in the bias session – as the selection of these non-preferred intentions required inhibiting the activation of the preferred intention representation.

2.3. 'Bias effect': between-experiment comparisons

2.3.1. Analyses

Finally, some analyses were conducted to directly assess whether the contribution of prior knowledge to the inference of an intention depended upon the target (motor vs. superordinate) and the scope (non-social vs. social) of the intention. Student tests were first conducted on the overall performance in the bias session between the motor experiments and the superordinate ones, and between the non-social experiments and the social ones. Second, the bias effects for each dimension (scope and target) were compared with each other. A score reflecting that effect for each type of intention (motor, superordinate, social, non-social intention) was calculated by subtracting, in each experiment, participants' hit rate for the preferred intention to the hit rate for the non-preferred intentions. The resulting scores were then entered in a 2 (motor vs. superordinate) x 2 (social vs. non-social) x 3 (amounts of information) factorial ANOVA.

2.3.2. Results

Contrasting the overall performance between experiments revealed no significant differences, showing that overall participants performed at comparable levels across the four experiments. Comparing the bias effect between experiments revealed significant effects of both the scope and the target of the intention. The bias effect was indeed significantly increased for the superordinate intention compared with the motor intention ($F(1,116)=8.36$, $p<.005$, $\eta^2 = .81$) and for the social intention compared with the non-social intention ($F(1,116)=5.06$, $p=.02$, $\eta^2 = .61$). Furthermore, these differences were observed for

different amounts of information according to the dimension itself. Indeed, along the scope dimension, the bias had a significantly greater effect when inferring a superordinate intention than a motor one in the conditions with a moderate amount of visual information (post-hoc Fisher test, $p = .014$). Along the target dimension, on the other hand, the only significant difference was observed for a high amount of visual information with a greater bias effect for inferring social intentions than non-social ones (post-hoc Fisher test, $p = .0035$) (see figure 8).

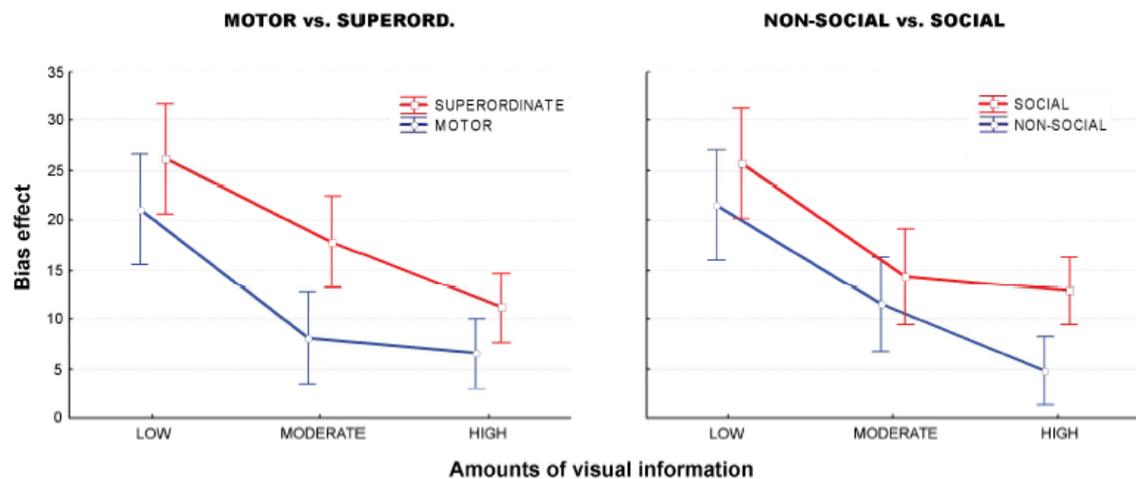


Fig. 8. Mean score (\pm SD) of the bias effect expressed as a percentage of correct responses for each type of intention. **Left panel:** comparison between intentions with same target but different scopes (MOTOR vs. SUPERORD.). **Right panel:** comparison between intentions with same scope but different targets (NON-SOCIAL vs. SOCIAL).

2.3.3. Preliminary discussion

Comparing the bias effect across the four experiments, two main results emerged. First, consistent with the previous results, we found the bias to differentially affect performance according to the scope and the target of the intention. The bias effect was indeed significantly more important in superordinate conditions than in motor ones and in social conditions than in non-social ones. Second, this effect varied according to the amount of visual information available to the participants. It was significantly more important for superordinate intentions than for motor intentions when the amount of visual information was moderate and greater in the condition of high amount of information when the intention occurred in a social context, rather than in isolation. These results show that for a low amount of visual information, the prior expectations equally contributed to the inference of any type of intention. Yet, as the amount of visual information increased, a differential contribution of these expectations was observed according to the dimension of the intention. Together, these results confirm those we previously obtained. Crucially, they suggest that participants' prior expectations contribute in a different way to the inference process depending on the type of intention that is focused on. We further consider these differences in the final discussion of the paper.

3. General discussion

The present study aimed at investigating how two distinct sources of information – perceptual (bottom-up) evidence and *prior* (top-down) expectations – interact to enable

one to make an intentional inference. To do so, we manipulated the participants' prior expectations about the probability of the underlying intention while varying the amount of visual information of the action scene. Our second purpose was to determine whether the contribution of these two sources of information would vary depending on the scope (motor vs. superordinate intention) and the target of the intention (social vs. non-social intention) that had to be inferred. To test this second hypothesis, we therefore manipulated the type of intention underlying the observed action using four distinct tasks (motor non-social, superordinate non-social, motor social and superordinate social tasks).

Two main results emerged. First, we observed that the intentional judgment indeed rests on an interplay of the participants' prior expectations (their probability of being true varied across the blocks) and of the reliability of sensory information available from the action scene. When this reliability decreased, the bias effect (i.e. the contribution of prior expectations) on performance increased, with participants responding more towards intentions they estimated as being the most likely cause of the observed behaviour. Second, this interaction was found to vary according to the *type* of intention, defined here by its scope (motor vs. superordinate) or its target (social vs. non-social). Indeed, directly comparing performance between intentions of different scopes but identical targets, and between intentions with the same scope but distinct targets, revealed an increase in the bias effect for both superordinate and social intentions. While this effect was only observed when the amount of visual information was low in the motor task, it was found to be significant for both low and moderate amounts of information in the superordinate task, and for any amounts of visual information in the social conditions.

Taken together, these results indicate that the degree of participants' prior knowledge contribution is sensitive to the type of intention that is focused on. From motor to superordinate, and from non-social to social intentions, participants' prior expectations exerted an increasing influence on their responses, at the expense of the sensory information available from the action scene.

3.1. Interaction between perceptual and prior information

In the 4 experiments of the present study, the degree of prior expectations contribution strongly depended on the reliability of the visual information conveyed by the video scenes. In conditions with low amount of visual information, whatever the type of intention, participants tended to give priority to preferred intentions at the expense of non-preferred intentions ; that is, to rely mostly on the intention they estimated to be the most likely cause of what was observed. This tendency to favor prior knowledge over perceptual information may further be accounted for by considering intentional inferences as an *inverse problem* (Kilner et al., 2007a, 2007b; Baker et al., 2007; see also Csibra & Gergely, 2007). Inverse problems characterise situations in which the same sensory input can have different causes. This type of problem is commonly encountered in ambiguous perceptual tasks – such as those using bi-stable or degraded stimuli – the resolution of which requires appealing to prior knowledge or making further assumptions about the nature of the observed phenomenon (Mamassian & Goutcher 2001). The significant contribution of prior expectations in conditions of high visual ambiguity precisely suggests that when sensory information was not sufficient to unambiguously infer one intention, participants compensated by massively appealing to their prior knowledge (i.e. about the space of the agent's possible intentions). This strategy resulted in a significant shift towards the intention with the highest probability.

Overall, this result comforts the idea that, in situations of sparse or incomplete data, a successful inference depends on an adaptive integration between bottom-up information

(from the observation of behaviour) and top-down prior knowledge about goals or intentions (Baker et al., 2006). This integration is consistent with a mechanism complementing the available perceptual information when it does not sufficiently constrain the number of potential solutions, namely, the many competing intentions congruent with what is observed. In line with this assumption, some authors have suggested that inferring another person's intention necessarily requires sensorimotor information to be complemented with information about mental states and attitudes (de Lange et al., 2008). Besides, it has been demonstrated that prior expectations are frequently used by children, even at a very early age. This tendency combines with a tendency to interpret actions as being directed towards a goal ('teleological obsession', Csibra & Gergely, 2007). When the visual information is not sufficient for interpreting the action as a goal directed one (Csibra, Gergely, Bíró, Koós, & Brockbank, 1999), or when the action is incomplete (Onishi, Baillargeon, & Leslie, 2007), children posit states of the world occasionally counterfactual to the perceptual evidence (such as the presence of occluded physical objects). The results of the present study are consistent with the existence of such a mechanism of data completion/correction operating through the default use of prior expectations. Crucially, however, we further show that reliance on this mechanism also depends on the type of the intention to be inferred, according to its scope (motor vs. superordinate) or its target (non-social vs. social).

3.2. Motor vs. superordinate intentions

Both non-social motor and superordinate experiments required recognising one motor act, the superordinate condition also requiring the final goal of the sequence (i.e. the shape being constructed) to be taken into account. Yet, across both experiments, prior expectations were found to differentially contribute to the participants' response. In the motor non-social experiment, a bias response towards the preferred intention was only observed in the condition where the amount of visual information was low. When participants were exposed to a moderate amount of visual information, these expectations no longer exerted an influence on performance, which then substantially depended upon the processing of the visual information alone. As previously suggested, this result may be explained by the fact that motor intention stands to action in a one-to-one relation: motor intentions like 'transport', 'rotate' or 'lift' a cube are directly accessible to the viewer from mere observation of the motor acts. On the other hand, a heightened contribution of prior knowledge is observed when shifting from a relation of strict correspondence (one-to-one relation) to a relation where the current action only represents one step, among others, of the final intention. In the superordinate experiments, prior knowledge was indeed found to significantly influence participants' performance in conditions of both low and moderate amounts of visual information.

One explanation for this increase in response bias might reside in the differences in complexity between both tasks. Indeed, participants were required to attend to a single motor act in the motor conditions while the superordinate conditions involved paying attention to a sequence of three successive motor acts. In superordinate conditions, this supplementary difficulty could have led subjects to adopt a strategy of the least effort that would have consisted in disregarding the viewed sequence and relying primarily on their prior knowledge. However, if that was the case, poorer performances should have been observed in these conditions. Yet, participants performed at comparable levels across motor and superordinate experiments showing that the differences in terms of contribution of prior information between both experiments are accounted for by the type of intention being considered rather than by the difficulty of the task.

Contrary to motor intention, superordinate intention stands to action in a many-to-one relation, meaning that the very same intention can be achieved by several distinct (commutative) sequences of actions. Because of this unpredictability, participants were thus unable to infer the underlying intention on the basis of the sole visual information arising from the first two actions. However, the present results suggest that, despite the unpredictability of the sequence, participants still initiated a response, before observing the last action, by appealing massively to their prior expectations.

This early use of prior expectations may be accounted for by the existence of a system that would pre-process the current action chain depending on the sequences previously encountered. Observing the beginning of an action, or a sequence of actions, would automatically activate a representation of the preferred intention that would be progressively suppressed or reinforced as the amount of visual information increases. Such a pre-processing would be particularly salient in superordinate conditions, where the beginning of the act chain proved to be of little importance for inferring the final intention it achieved. As such, it would explain why selecting a non-preferred intention in bias sessions induced a significant cost on participants' performance. In these sessions, selecting a non-preferred intention would indeed imply disengaging from the early activation of a preferred and better represented intention. Finally, such pre-processing may account for why prior expectations are favoured over visual information in conditions of high perceptual uncertainty, as it would account for the role that priors continue to play when the amount of perceptual information increases. In superordinate conditions, the current sequence of actions would pre-activate the representation of the preferred intention (i.e. the intention with the highest probability) to such an extent that a greater amount of visual information would be required to counteract it.

3.3. Non-social vs. social intentions

Social experiments were characterized by participants' response over-relying on prior expectations as revealed by responses massively shifting towards preferred intentions (i.e. "tit-for-tat" intention (TFT): cooperation if previous cooperation, defection if previous defection) whatever the amount of visual information available from the action scene. This increased reliance on prior knowledge cannot be accounted for by differences in terms of complexity between non-social and social experiments, namely a greater memory load due to the requirement to track two successive intentions – the first and the second player's ones. Indeed, participants performed equally well, in terms of correct responses and reaction times, in both the social and non-social experiments. Additionally, the effect of facilitation associated with TFT strategy in the motor social experiment cannot be explained by a visual priming effect of the first player's action on the second player's one, that could have occurred when the latter performed the same action as the former. Indeed, TFT strategy was also found to be favoured in the superordinate social experiment; yet in this study TFT strategy did not necessarily imply that the action of the first player should be reproduced by the second player.

The dependence of the participants on their prior knowledge appears to reflect some expectations driven by the social context of the task. It is well-known that even basic movements, like the relative movements of geometrical figures, automatically induce participants to perceive the figures as socially interacting (Heider & Simmel, 1944; Castelli, Happé, Frith & Frith, 2000; Scholl & Tremoulet, 2000, for a review), together with eliciting strong expectations about the intentional causes of their movements (striking, kissing, escaping, etc.). Situations identified as involving social interactions are generally prone to trigger specific expectations concerning the way agents are likely to behave in such

situations (Frith & Frith, 2006). These expectations may be derived from knowledge of diverse nature, such as that provided by group stereotypes (Fiske, Cuddy, Glick, & Xu, 2002), social-specific naive theories (Ybarra, 2002) or agent's reputation acquired from experience of reciprocal social interactions (Singer et al., 2006). In the present experiment, increased dependence on prior knowledge for inferring social intentions, regardless of their scope, seems precisely to fall within expectations of this kind. Indeed, during the whole task, increasing the frequency of the second player adopting a TFT strategy amounted to progressively assigning a specific reputation to that player. A bias in the response towards "tit-for-tat" mode of reciprocity reveals that participants did integrate this reputational knowledge and made their response accordingly.

The pervasive effect of these specific expectations is also well illustrated in the baseline condition by the early preference of participants for the TFT reciprocation. Even before being biased in this direction, participants tended to infer more rapidly that the second player was more inclined to mirror the first player's strategy. This early preference was probabilistically reinforced in the bias session and, as a consequence, exerted a greater influence on the participants' performance since it persisted even when the reliability of the visual information was high. Indeed, while in the motor non social experiment the very same motor act presented alone was inferred from a much lower amount of visual information, in the social experiment a bias response towards the preferred social intention was still observed for a higher amount of visual information. This shows that the influence of these expectations in the bias session was such that the participants had difficulties in disengaging from their *a priori* expectations, resulting in predicting a play congruent with prior expectations but counterfactual to perceptual evidence. Similarly, these difficulties could account for the cost in performance associated with the selection of intentions that did not meet these expectations. In the bias session, participants were indeed significantly less accurate and were slower to select a non-preferred intention (i.e. always defect, always cooperate) when this selection required concomitantly inhibiting the competing tit-for-tat intention.

3.4. Simulation vs. reasoning accounts of action understanding

The two main results of the present study (interaction between prior expectations and perceptual information and the modulation of this interaction as a function of the type of intention) may help reconcile the two major accounts of action understanding developed over the last decade. On the simulation account, we understand our conspecifics' intention by literally simulating their action via the activation of our own motor planning system; the result of this process of internal replication is the selection, in the observer's own repertoire, of the intention that would have caused the very same action. This type of explanation mostly stresses the role of sensory information derived from the kinematics of the action (Rizzolatti et al., 2004). By contrast, the "theory theory" account postulates that action understanding is based on specialized inferential processes and mostly emphasizes the contribution of the context-related prior knowledge derived from our intuitive theories of human behaviour, on the one hand, and on the subject's past experiences and rules she has drawn from them, on the other (Leslie, 1987; Gopnik & Wellman, 1992, 1994).

A wealth of empirical data and theoretical works nowadays converges on the idea that these two major classes of mechanisms play a complementary role in intention inference (Mitchell, 2005; Keysers & Gazzola, 2007; Brass, Schmitt, Spengler, & Gergely, 2007; de Lange et al., 2008). The results of the present study comfort these observations. By suggesting that intentional judgment relies on a relative balance of bottom-up sensory and

top-down prior information, they plead in favor of a *hybrid* model of action understanding. In such a model, the observer would mobilize either low-level simulation or high-order inferential mechanisms depending on whether the current sensory evidence is, or is not, reliable enough to elicit simulation from observation.

Recently, Kilner and colleagues proposed a theoretical framework that attempts to further account for how these two classes of mechanisms may interact to enable one's understanding of other people's intentions. This framework relies on the hierarchical architecture of action representations ranging from the intention level to the kinematics level (see also Grafton & Hamilton, 2007). In this architecture, the selection of one type of action representation would result from the resolution of the inverse problem at each level of the hierarchy. Basically, each level uses a model to generate a prediction of the representations in the level below. This prediction is then compared with the representation at the subordinate level and prediction errors arising from that comparison are returned to the higher level to adjust its representation. This adjustment is generalised to the different levels of the hierarchy (intention, motor command and kinematics). The most likely cause of the observed action is then inferred by minimising the prediction error at all the levels of this hierarchy (Kilner et al., 2007a, 2007b). Given visual kinematics, goal expectations are first generated, from these goal representations motor commands are then predicted and given these motor commands, kinematics are in turn predicted. In this framework, top-down influences are therefore dynamically generated since the estimates produced at the higher levels become prior expectations for the lower levels.

Our results can be consistently interpreted in the light of the Kilner's hierarchical model. A motor intention can be directly predicted from the observation of the current motor act, provided the related visual information is sufficient to enable comparison with expected kinematics at higher levels. In this case, participants' performance is strongly dependent on minimising the prediction error that arises from this comparison. However, this comparison also closely depends on the reliability of the current movement kinematics; when the amount of visual information is too low, this comparison cannot be made, and, as a result, subordinate levels cannot adjust their representation to higher estimates of the hierarchy. We observed that, when this comparison could not be carried out, participants consistently appealed to their prior knowledge. In a hierarchical model of action representations, such an over-reliance on priors could be made possible by the existence of a short circuitry of recursive loops between subordinate and higher levels of the cortical hierarchy. These recursive loops would be mobilized when data is sparse to shortcut the automatic comparison process between observed and expected kinematics movement. Importantly, the engagement of this mechanism proved to be dependent on the amount of visual information available from the action scene, but independent from the scope and target of the intention, since it was observed to operate at the lowest levels of visual information in each of the four experimental conditions.

Noteworthy, the engagement of these recursive loops is also sensitive to variations in the relationship between the observed action and its goal. Superordinate conditions indeed involved a more important recourse to participants' prior expectations even when the visual information significantly increased to a moderate (non-social) or even a high level (social). This greater dependence on prior expectations can be explained by the fact that, in superordinate conditions, many competing intentions are congruent with the visual information conveyed by the current motor act. Thus, whereas minimising the prediction error between expected and current kinematics may be sufficient to predict the agent's single act (e.g. to rotate), it may not be to infer unambiguously which of the multiple superordinate

intentions (e.g. final shapes) it contributes to accomplish. As a consequence, we found the weight of the decision to be mostly carried by participants' prior expectations, suggesting, in this situation of accrued perceptual uncertainty, an early shortcut of the comparison process between levels of the action representation hierarchy. Crucially, this shortcut was independent of the amount of information, since it occurred even when the visual information was high enough for the participant to be normally confident about what she is seeing. This observation suggests that recursive loops of this kind could be mostly recruited in contexts where relying on one's prior expectations is a better guarantee for accurate inference, even if such expectations can occasionally go against perceptual evidence.

4. Conclusion

Our results shed light on how sensory information, derived from the kinematics of the observed action, interacts with prior expectations to enable one's understanding of other people's intentions. We first showed that the contribution of participants' prior knowledge was sensitive to the reliability of the sensory information available from the action scene. A greater contribution of this knowledge was observed in condition of sparse visual information, suggesting the engagement of a mechanism of data completion operating through the default use of prior expectations.

Second, we found that the priors' contribution also depended on the type of intention that was inferred. An increased reliance on priors was indeed observed in conditions where the agent's intention could not be predicted by the sole visible, current motor act, but further required estimating the superordinate goal this act contributed to achieve. In this case, participants' expectations – being progressively acquired from observation – were found to most frequently supersede the visual information conveyed by the current motor kinematics. Thus, the more participants responded towards the biased (e.g. expected) intentions, the more the visual information tended to play a confirmatory, rather than a predictive role. Such a shift in the contribution of visual evidence is likely to account for why participants, in this condition, mostly over-relied on their priors to make their decision, even though it ran counter to the perceptual evidence.

Crucially, an over-reliance on priors was also massively observed in social conditions. We suggested that the early influence of social-specific expectations (e.g. expectations on how agents are the most likely to behave in a context of reciprocal interaction) may account for this important shift in the response toward participants' priors. Contexts of social interaction are indeed prone to elicit modular, high-level expectations, which may contribute to giving priority to some intentional causes (e.g. cooperation if previous cooperation, defection if previous defection) at the expense of other competing causes. These *a priori* expectations, being acquired from experience (probabilistic bias) or derived from domain-specific knowledge (TFT reciprocation), were found to favour some action representations so that less sensory evidence was needed for the participants to be confident about their decision, *i.e.* about which kind of intention was most likely the cause of the observed action.

Acknowledgments

This research was supported by a grant of the Agence Nationale de la Recherche (ANR-ZR81). V. Chambon and P. Domenech were supported by scholarships from the French Ministry of Research. This research was approved by the local Ethical Committee (B80631-60). The authors report no competing interests. We are grateful to Mathilde Vernet,

Juliette Sablier and Mariia Kaliuzhna for their help in preparing the stimuli. We also wish to thank Coralie Chevallier, Hugo Mercier, Nicolas Baumard, and Guillaume Barbalat for their helpful comments and suggestions on an earlier draft of this article.

Appendix

For each experiment, a pre-test was conducted in order to i) control for the temporal homogeneity of the action sequences, and ii) to select three temporal windows relevant for the subsequent experiments (i.e. Inf1, Inf2, Inf3 or low, moderate, high amount). All movies were filmed using a digital camera (Sony®- HDR-SR7) and were acquired and tailored using the software Adobe Premiere®.

A. Intra- and inter-sequence comparisons

Prior to build each experiment, a pre-test was conducted to control the temporal dynamics of each action sequence. Specifically, we controlled that, for any point within a given sequence, a particular type ('transport' vs. 'lift' vs. 'rotate') or model (transport[1] vs. transport[2], lift[1] vs. lift[2], etc.) of action was not more or less discriminable than another one. To do so, each action scene in each of the 4 experiments was first cut at {960, 1160, 1320, 1400, 1480, 1560, 1640, 1720, 1800, 1880, 1960, 2080} ms after the onset of the movement. The resulting movies (12 clips per model of action) were randomly presented to a group of 12 participants in four sessions. Each session consisted in a basic action recognition task. In the two first sessions, movies from the motor and superordinate experiments were presented with the instruction of recognizing the performed actions by pressing, as quickly and as accurately as possible, one of the corresponding keyboard buttons. In the two last sessions, movies from the social motor and social superordinate experiments were shown with the same instruction. The order of sessions was counterbalanced across participants. For the (non-social) superordinate and social (motor and superordinate) movies, the only last action was shown.

Two groups of statistical analyses were conducted on both hits and RTs.

First, we compared participants' hits for each type of action *within* each session ('lift' vs. 'rotate' vs. 'transport' actions within non-social motor and superordinate experiments; 'transport' vs. 'rotate' actions within social motor and superordinate experiments), then we compared hits for actions of the same type *between* sessions (e.g. lift[motor] vs. lift[superord.]).

Two-tailed t-tests revealed no significant differences between participants' correct responses for the different types of action within each session (comparing 'transport' vs. 'lift' vs. 'rotate' actions in non-social motor and superordinate conditions, all $p > 0.22$; comparing 'transport' vs. 'rotate' actions in social motor and superordinate conditions, all $p > 0.4$), as well as no significant differences for actions of the same type between sessions (non-social conditions: lift[m] vs. lift[s], rotate[m] vs. rotate[s], transport[m] vs. transport[s]: all $p > .19$; social conditions: transport[m] vs. transport[s], rotate [m] vs. rotate[s]: all $p > .25$).

Second, we were compared the individual distributions of hits and reaction times (RTs) across the different movie segments to ensure that participants' performance was equally sensitive to variations in the amounts of visual information (see figure 9). To do so, we computed within each experiment a one-way Kruskal-Wallis test with participants' hits and RTs as dependent variables and the amount of information as a 12-level factor.

As shown in Figure 10, the participants' detection curve showed the usual sigmoidal shape, performance rapidly increasing within a 1400-1700 ms interval of movie durations up to the maximal value (100%). In all experiments, Kruskal-Wallis tests revealed no significant differences between individual distributions of performance, with, between participants, identically decreasing RTs (motor: $H(11,144)=9.34, p=.59$; superordinate: $H(11,144)=7.64, p=.74$; social motor: $H(11,144)=6.34, p=.84$; social superord.: $H(11,144)=5.1, p=.89$) and increasing rate of correct responses (motor: $H(11,144)=2.88, p=.9$; superordinate: $H(11,144)=2.61, p=.9$; social motor: $H(11,144)=1.98, p=.9$; social superord.: $H(11,144)=1.02, p=.9$) as the available amount of visual information increased.

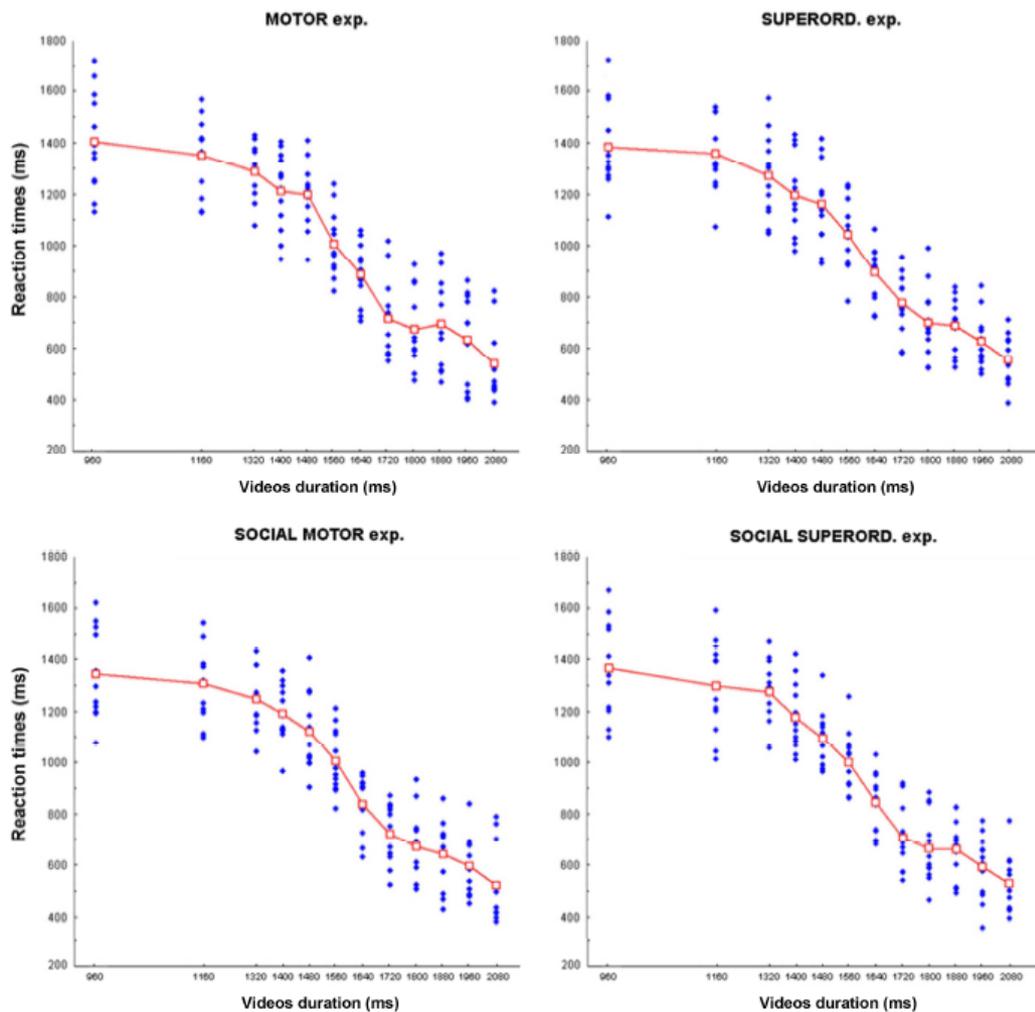


Fig. 9. Distribution of participant's reaction times (blue dots) across the 12 movie segments. Reaction times for the different actions were pooled across subjects. Red squares: mean reaction times across participants for each of the 12 duration ranges.

B. Selection of low, moderate and high amounts of information

For the needs of the subsequent experiments (social and non-social motor and superordinate exp.), 3 different amounts of visual information (corresponding to three different movie segments) were selected among the 12 controlled ones. To do so, a Weibull psychometric function was fit with the restricted least squares method to the data (we used a Weibull function because such function generally provides a good model for contrast discrimination and detection data ; see Nachmias, 1981; Wichmann & Hill, 2001). We then computed the amount of visual information corresponding to the inflexion point of the sigmoid curve within each experiment (shown as a blue dot on the resulting regression curves, see figure 10). The duration corresponding to the inflexion point(1560ms –moderate) and the two nearest durations (1480 –, low and 1640 ms – high) were selected for the covert blocks. This choice ensured equal timing steps as well as similar performance differences between the three levels. The 1880 ms duration ($\approx 100\%$ hit rate) was chosen for the 'overt' trials.

The Weibull function is described by:

$$F(x, \alpha, \beta, \gamma) = 1 - \gamma \exp \left[- \left(\frac{x}{\alpha} \right)^\beta \right], \quad 0 \leq x < \infty.$$

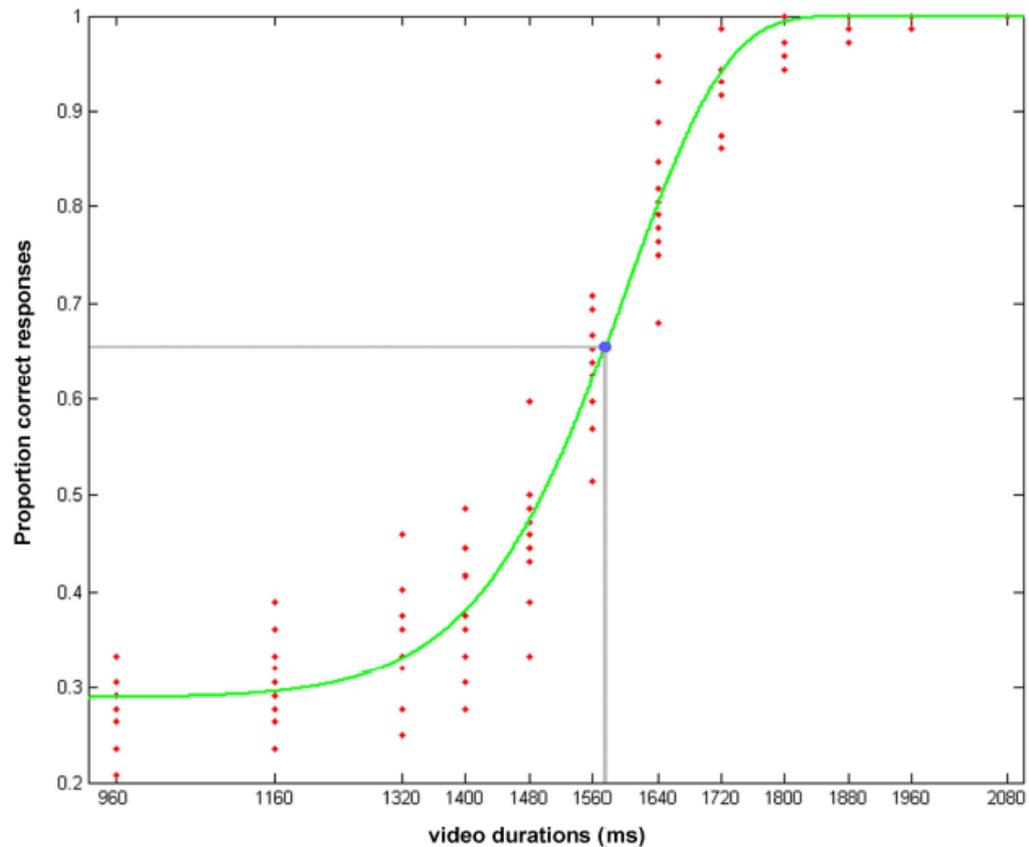


Fig. 10 . Motor experiment: psychometric curve (green) fit to the cumulative distribution of participant's correct responses (red dots) across the 12 movies segments. Responses for the different actions were pooled across. The blue dot refers to the inflexion point of the sigmoid curve. In each experiment, the inflexion point occurs at the following duration: **A. Motor:** 1576 ms. **B. Superord.:** 1558 ms. **C. Social motor:** 1546 ms. **D. Social superord.:** 1550 ms.

References

- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, 11, 1109-16.
- André, J.B., & Day, T. (2007). Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. *J Theor Biol*, 247, 11-22.
- Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition*. Princeton, Princeton University Press.

- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2006). Bayesian Models of Human Action Understanding. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18, pp. 99-106), MIT Press.
- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2007). Goal Inference as Inverse Planning, In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*.
- Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention, *Consciousness and Cognition*, 17, 557-64.
- Blakemore, S. J., & Decéty, J. (2001). From the perception of action to the understanding of intention. *Nature Neuroscience*, 2, 561-567.
- Blakemore S. J., & Frith, U. (2004). How does the brain deal with the social world? *Neuroreport*, 15, 119-28.
- Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes vs. action simulation. *Current Biology*, 17, 2117-2121.
- Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.
- Carey, S. (1985). *Conceptual Change in Childhood*, Cambridge, MA: MIT Press.
- Castelli, F., Happé, F., Frith, U. & Frith, C. D. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314-325.
- Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., & Walter, H. (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*, 45, 3105-13.
- Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, 26, 237-67.
- Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychol*, 124, 60-78.
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering H (2008). Complementary systems for understanding action intentions. *Current Biology*, 25, 454-7.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J Pers Soc Psychol*, 82, 878-902.
- Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, 1079, 36-46.
- Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, 17, 724-32.
- Fu, G., & Lee, K. (2007). Social grooming in the kindergarten: the emergence of flattery behavior. *Developmental Science*, 10, 255-265.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 596-609.
- Gallese, V., & Goldmann, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci*, 2, 493-501.

- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci*, 8, 396-403.
- Georgiou, I., Becchio, C., Glover, S., & Castiello, U. (2007). Different action patterns for cooperative and competitive behaviour. *Cognition*, 102, 415-433.
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36, 299-308.
- Goldman, A. I., & Sebanz, N. (2005). Simulation, mirroring, and a different argument from error. *Trends Cogn Sci*, 9, 320.
- Gopnik, A., & Wellman, H. (1994). The theory theory. In Hirschfeld, L. and Gelman, S. (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge: Cambridge University Press.
- Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, 7, 145-71.
- Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representation in humans by positron emission tomography: 2. Observation compared with imagination. *Experimental Brain Research*, 112, 103-111.
- Grafton, S. T., & Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum Mov Sci*, 26, 590-616.
- Grèzes, J., & Decéty, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, 12, 1-19.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition, In Ron Sun (ed.), *The Cambridge handbook of computational cognitive modelling*, Cambridge: Cambridge University Press.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behaviour. *American Journal of Psychology*, 57, 243-259
- Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends Cogn Sci*, 9, 21-25.
- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language*, 23, 190-223.
- Jacob, P. (2009). The tuning-fork model of human social cognition: a critique. *Consciousness and Cognition*, 18, 229-43.
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, 14, 103-109.
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation, and self-identification. *Mind and Language*, 19, 113-146.
- Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends Cogn Sci*, 11, 194-6.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). Predictive coding: an account of the mirror neuron system. *Cogn Process*, 8, 159-66.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, 18, 619-23.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308, 78-83.

- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philos Trans R Soc B*, 363, 2021-31.
- Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn Sci*, 10, 319-26.
- Leslie, A. M. (1987). Pretense and representation: the origins of "theory of mind". *Psychologica I Review*, 94, 412-42
- Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, 81, B1-9.
- Mele, A. (1992). *Springs of Action: Understanding Intentional Behavior*, Oxford University Press.
- Mitchell, J. P. (2005). The false dichotomy between simulation and theory-theory: the argument's error. *Trends Cogn Sci*, 9, 363-364.
- Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research*, 21, 215-223.
- Nichols, S., & Stich, S. (2003). *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds*, Oxford University Press.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychol*, 124, 106-28.
- Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, 22, 129-151.
- Pacherie, E. (2000). The content of intentions. *Mind and Language*, 15, 400-432.
- Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition*, 107, 179-217.
- Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*, University of Michigan Press.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Neurosci Rev*, 2, 661-670.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu Rev Neurosci*, 27, 169-92.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn Sci*, 10, 70-6.
- Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: the role of expertise. *Psycho Bull Rev*, 16, 70-5.
- Saxe, R. (2005). Against simulation: the argument from error. *Trends Cogn Sci*, 9, 174-9.
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends Cogn Sci*, 4, 299-309.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press.
- Searle, J. (1995). *The Construction of Social Reality*, New York: The Free Press.

- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466-9.
- Ybarra, O. (2002). Naive causal understanding of valenced behaviors and its implications for social information processing. *Psychol Bull*, 128, 421-41.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293-1313.

- Etude chez le patient schizophrène - Mindreading under influence: abnormal dependence on prior knowledge in patients with schizophrenia

Valerian Chambon, Guillaume Barbalat, Pierre Jacquet, Elisabeth Pacherie, Nicolas Franck, and Chloé Farrer

(preliminary draft)

1. Introduction

Schizophrenia is associated with poor social functioning, as shown by impairments that most patients exhibit in interpersonal communication and relationships (Corcoran, 2001, for review). To account for these deficits, Theory of Mind (ToM) function – the ability to explain and predict others' behaviour on the basis of their mental states – has often been focused as a probable candidate (Frith, 2004; Harrington et al., 2005; Sprong et al., 2007). Though ToM dysfunction in schizophrenia has given rise to an extensive line of research these two last decades, both its nature and the extent to which it is present are still widely debated (McCabe et al., 2004; Zalla et al., 2006; Walter et al., 2009). The reason for this controversy may stem from the fact that, taken together, the existing data proves to be particularly contradictory. Most studies indeed failed to formalize coherent associations between symptom dimensions and patients' defective abilities in intention attribution and mentalizing. Thus, while a number of authors report selective correlations between a mentalizing deficit and the positive dimension of the disease – such as formal thought and language disorder (Sarfati et al., 1999; Brunet et al., 2003), or reality distortion (Frith & Corcoran, 1996) –, some other studies found concurrent associations with either disorganization (Zalla et al., 2006) or psychomotor poverty syndroms (Mazza et al., 2001), or with negative behavioural signs (Pickup & Firth, 2001) such as flat affects, social withdrawal and alogia (Langdon et al., 2002).

It has been suggested that global inconsistency between patients' symptoms and ToM dysfunction may be rooted in the great variety of the tasks proposed, both in terms of complexity and type of (verbal, iconographic) material being used (see Walter et al., 2009). Crucially, the heterogeneity of the data could also result from the lack of control of the notion being tested, namely, the other's intention the patient has to infer. "Intention" is indeed an umbrella term that may embrace various sub-types each being assigned different scopes (*i.e.* proximal or distal goals) and targets (*i.e.* action directed at an object or at another agent) (Pacherie, 2000, 2008). The ability to appreciate other people's intentions thus refers to separate processes which could be differentially impaired depending on the scope and/or

the target of the intention represented. As such, one cannot preclude the possibility that the schizophrenic patients' deficit in mentalizing may embrace one particular *type* of intention while sparing appreciation of the others. It then appears to be of an extreme importance to precisely assess at which (possible) level of intentional complexity this deficit might take place, in strictly controlling the type of intention being evaluated on the basis of a careful intentional typology.

To do so, we adapted from Chambon et al. (xxxx) a series of tasks that require taking into account four distinct types of intentions in a large sample of healthy controls and symptom-matched patients. Both groups were instructed to identify, from observation of videos of action sequences, intentions that varied depending on either their scope (motor vs. superordinate) or target (social vs. non-social). Thus, combining each of these two dimensions resulted in 4 distinct experiments, in which participants were successively required to infer

i) motor (proximal) intentions, *i.e.* intentions directed at simple motor goals, such as lifting an arm, pressing a button, reaching for an object, etc. (**task 1**)

ii) superordinate (distal) intentions, *i.e.* intentions directed at somewhat complex or general goals, the achievement of which typically involves the completion of a number of subgoals or substeps (*e.g.* drinking a glass of wine: taking a glass, opening a bottle, pouring the wine into the glass, etc.) (**task 2**)

iii) social motor intentions, *i.e.* simple goals achieved in a context of social interaction with another agent (**task 3**)

iv) social superordinate intentions, *i.e.* general goals comprising a sequence of basic motor acts achieved in a context of social interaction (**task 4**)

So far, only few studies directly tested the patients' ability to appreciate distinct types of intention within the same experimental settings, or using the same material across conditions. Among them, evidence revealed that schizophrenic disorganised patients show a specific impairment in evaluating distal non-social intentions underlying current action sequences, while their ability for inferring proximal intentions would be intact (Zalla et al., 2006). Incidentally, this impaired ability would arise together with executive disorders in action sequencing, that is, with difficulties to coordinate the different segments that compose the action in relation with the mental states of the protagonists (Zalla et al., 2004; Zalla et al., 2006). However, a recent study manipulating non-social and social intentions suggested that patients could be impaired, not in appreciating actions directed at inanimate objects, but specifically in inferring intentions achieved in a context of (present or forthcoming) social interaction (Walter et al., 2009). A pronounced inability to appreciate social intentions was further associated with an abnormal modulation of activity in regions typically activated in ToM tasks, *i.e.* paracingulate cortex and bilateral temporo-parietal junctions (Walter et al., 2009). Interestingly, this reduced activity for social intentions was partly the consequence of no signal drop in physical (non-intentional) conditions, revealing that patients' mentalizing profile was also characterized by a significant tendency to (over-)attribute intentionality to inanimate objects (see also the 'hyper-TOM' assumption, Abu-Akel & Bailey, 2000).

In the present study, we hypothesized that this confusing array of findings may be disentangled by investigating patients' mentalizing abilities at a more fine-grained level of meta-cognitive functioning; that is, not only by assessing patients' raw performances in intention recognition, but in further exploring how individuals with schizophrenia deal with the information which usually contributes to that recognition. We indeed showed in a previous study that inferring other's intentions requires taking into account two distinct

(external and internal) sources of information: *i*) first, the *sensory information* available from the action scene and derived from the agent's movement kinematics, and *ii*) second, the observer's *prior expectations* about which intention is, given past experience, the most likely cause of what is observed. Intentional inference can be modelled as an adaptive interaction between these two sources of information (see Baker et al., 2006), with participants tending to rely progressively more on their priors as the reliability of sensory information decreases, and conversely (Chambon et al., xxxx ; see also Onishi et al., 2007). Converging evidence indicates that difficulties in action understanding observed in schizophrenia might further reflect an abnormal over-reliance on either of these two types of information, depending on the patients' prevalent symptoms. As an example, the mentalizing profile of productive patients may be characterized by a tendency to give excessive credit to endogenous, self-generated information – e.g. prior expectations of how people are supposed to behave under some circumstances – whereas negative patients would display a stimuli-induced mentalizing style that may be accounted for by an exaggerated tendency to focus on directly observable, external information, rather than inner experiences (Frith, 1992; Taylor, 1994).

In addition to assessing patients' abilities of drawing inferences at each level of intentional complexity, we sought to test this hypothesis by varying the following within each of the 4 above-mentioned tasks

(1) the amount of visual information available from the action scenes. We proceeded by manipulating the degree of completeness of the action sequences, resulting in video clips of varying durations

(2) prior expectations that participants had about the possible intentions of the observed agent. To do so, we progressively increased the probability that one intention (termed *biased* or *preferred intention*) occurs over the sequence, at the expense of other intentions (non-biased or *non-preferred intentions*) with the same scope and target.

We then looked at whether, for each of the four intentions considered, patients' inferences may be accounted for by an abnormal dependence on their priors and/or sensory evidence conveyed by the action scene, and whether over-relying on either of these sources of information did correlate with the severity of positive (SAPS; Andreasen, 1984), negative (SANS; Andreasen, 1983), or disorganisation symptoms of schizophrenia.

Overview of the present study

The study consisted of four intention recognition tasks, each combining one type of target (social vs. non-social intentions) with one kind of scope (motor vs. superordinate intentions). In each of the 4 tasks, participants were instructed to watch video sequences of intentional actions performed by either a single actor in isolation (non-social tasks) or by two interacting agents (social tasks), and to identify the intention guiding the observed action. As previously mentioned, two factors were varied over the sequences: (1) the visual information conveyed by the actor's kinematics, and (2) expectations about the most likely intention that participants had prior to seeing the action scene.

Regarding the first factor, 4 amounts of visual information have been selected (see Chambon et al., xxxx, Appendix Part A, for the selection and control of these amounts) : very high, high, moderate, and low, corresponding to varying films durations, ranging from {1880, 1640, 1560, 1480} ms after the movement onset, respectively.

Regarding the second factor, each experiment was divided into two sessions: a **baseline** session in which all intentions had the same probability to be achieved, and a **bias** session in which the probability of one particular intention (*preferred* intention) was

increased (55%) at the expense of the others (22 % each). Each session was in turn divided into two interleaved phases:

1. a learning phase (overt blocks) in which the action scenes were shown with a very high amount of visual information, and equal (baseline) or varying (bias session) probabilities of occurrence for each intention represented.
2. A test phase (covert blocks) which allowed one to assess the strength of the current bias by varying (low, moderate, and high) amounts of visual information. In this phase, each intention had equal probability to occur (see figure 1).

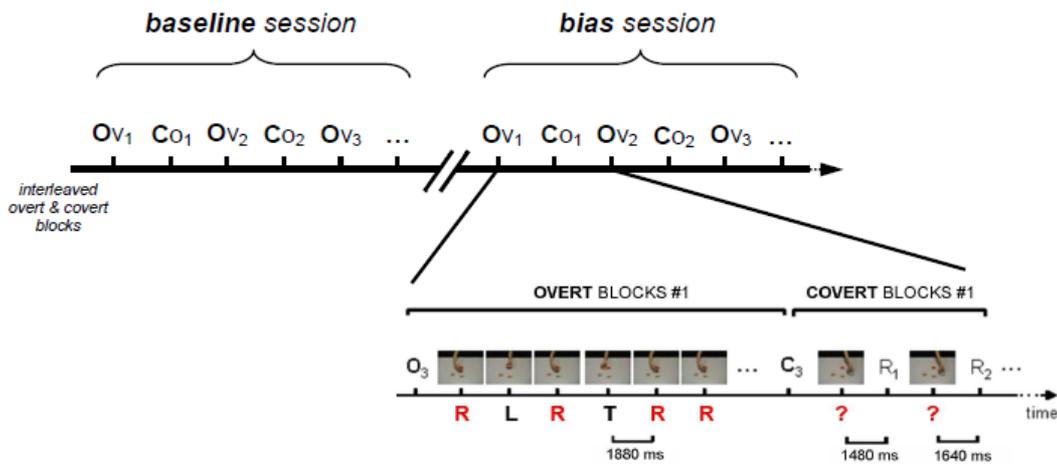


Fig. 1. Experimental design: Examples of a typical experimental sequence (one overt block followed by one covert block) used in both the baseline and the bias sessions.

Overt blocks: movies with a very high and constant amount of visual information. Covert blocks: movies with lower and varying amounts of visual information. In the 4 experiments, the probability of all intentions was held equal across the block, except in overt blocks of the bias session, where one particular intention had a greater probability to be accomplished than the other ones (here, the “rotate” action in red). Labels: Ov: Overt blocks; Co: Covert blocks; L: “lift” action; R: “rotate” action; T: “transport” action.

All movies were filmed using a digital camera (Sony®- HDR-SR7) and were acquired and tailored using the software Adobe Premiere®. The movies were presented on a computer monitor (IIYAMA® 19”) at a distance of 60 cm from the participant. Finally, tasks modeling intentions with the same scope or target were controlled to have close visual properties in order to allow direct comparisons between them.

Material and procedure were the same as those used in Chambon et al. (xxxx).

2. Method

2.1. Participants

Four groups of 20 patients with schizophrenia each and four groups of 30 healthy participants each volunteered to participate in the study. To each task (motor, superordinate, social motor, and social superord. tasks) were assigned one group of patients and one group of comparison participants among the aforementioned ones. Healthy participants

were systematically matched with patients for sex, age, handedness (Oldfield, 1971) and years of education (see Table 1). None of them reported psychiatric problems.

Patients recruited fulfilled DSM-IV (American Psychiatric Association, 1994) criteria of schizophrenia, with no other psychiatric diagnosis on DSM-IV Axis I. All patients were receiving antipsychotic medication and were clinically stable at the time of testing (Table 1). Negative and positive symptoms were evaluated with the SANS (Andreasen, 1983) and the SAPS (Andreasen, 1984). A disorganization score was also computed by summing the following subscores: bizarre behavior, positive formal thought disorder, alogia, and inappropriate affect. Each group of patients was matched with the three others for the severity of negative, positive and disorganization symptoms (Table 1). In the social motor task, one patient performed at the chance-level and was therefore excluded from analyses (n=19).

For all groups, exclusion criteria included history of neurological illness or trauma, alcohol or drug dependence according to DSM-IV criteria, analphabetism, and age older than 60 years. All participants reported normal or corrected-to-normal visual acuity. After complete description of the study to the subjects, written informed consent was obtained. This research was approved by the local Ethical Committee (B80631-60). All participants received 10 euros for taking part in the study.

CHARACTERISTICS	Age (years)	Educ. (years)	Handedness	Duration of illness (years)	SANS score ^a	SAPS score ^b	Disorg. score ^c
EXPERIMENT							
MOTOR							
Healthy (n=30)	35.1 (7.5)	11.9 (2)	0.87 (0.14)	–	–	–	–
Patients (n=20)	34 (9.3)	11.1 (1.7)	0.83 (0.16)	10.3 (7.5)	40.5 (15)	31.9 (23.5)	16.1 (12.9)
P value	.65	.12	.37				
SUPERORDINATE							
Healthy (n=30)	36.5 (8.9)	12.1 (1.5)	0.81 (0.17)	–	–	–	–
Patients (n=20)	34.6 (8.8)	11.6 (1.8)	0.78 (0.17)	12.3 (8.1)	43.2 (21.6)	29.9 (15.6)	12.9 (5.3)
P value	.46	.26	.61				
SOCIAL MOTOR							
Healthy (n=30)	34.2 (10.5)	11.4 (1.8)	0.82 (0.14)	–	–	–	–
Patients (n=19)	35.2 (9)	11.2 (1.7)	0.79 (0.19)	11 (8.4)	44 (24.1)	28.5 (22.3)	14.5 (12.8)
P value	.74	.61	.57				
SOCIAL SUPERORD.							
Healthy (n=30)	35.4 (8.8)	12.3 (1.9)	0.85 (0.13)	–	–	–	–
Patients (n=20)	33.8 (10)	11.7 (1.4)	0.81 (0.18)	11.9 (8.6)	44.8 (23.9)	29.4 (15.1)	11.5 (6.9)
P value	.56	.27	.48				
				<i>all p>.43</i>	<i>all p>.5</i>	<i>all p>.64</i>	<i>all p>.14</i>

^a SANS, Scale for the Assessment for the Negative Symptoms

^b SAPS, Scale for the Assessment of Positive Symptoms

^c Sum of the scores for bizarre behaviour, positive formal thought disorder from the SAPS, and alogia and inappropriate affect from the SANS

Table 1. Clinical and demographic characteristics (mean ± S.D.)

2.2. Non-social experiments

In the first experiment, participants were instructed to infer the *motor intention* (to lift, to rotate or to transport) of an actor manipulating meaningless objects (figure 2, **A**). In the second experiment, participants were instructed to infer the *superordinate intention* (i.e. the general goal) underlying a sequence of motor acts (figure 2, **B**).

2.2.1. Material

In both non-social experiments, visual stimuli presented were incomplete movies representing an actor's hand manipulating meaningless objects. The objects used were three rectangular cubes of similar size (3cm x 6cm) and orientation, placed at equal distance (16.8 cm) from the starting position of the actor's hand. Actions were performed on one single cube (motor experiment) or on each of the three cubes (superordinate experiment) in the direction of the centre of the scene. Each trial was unique, meaning that all the movies

were presented only once to prevent any influence of memorized kinematics parameters on participants' performance.

2.2.2. Procedure

A. Motor experiment: Each trial started with the onset of an incomplete movie showing an actor performing a basic action on one of the three available cubes. As soon as the sequence ended, participants were instructed to infer the intention underlying the actor's action by pressing, as quickly and as accurately as possible, one of the three corresponding keyboard buttons (T for 'transport', L for 'lift', or R for 'rotate'). In the bias session, the type of motor intention the probability of which was increased (*i.e. preferred* intention) was counterbalanced across participants. Finally, each action was performed equally often with all objects (cube 1, 2, or 3).

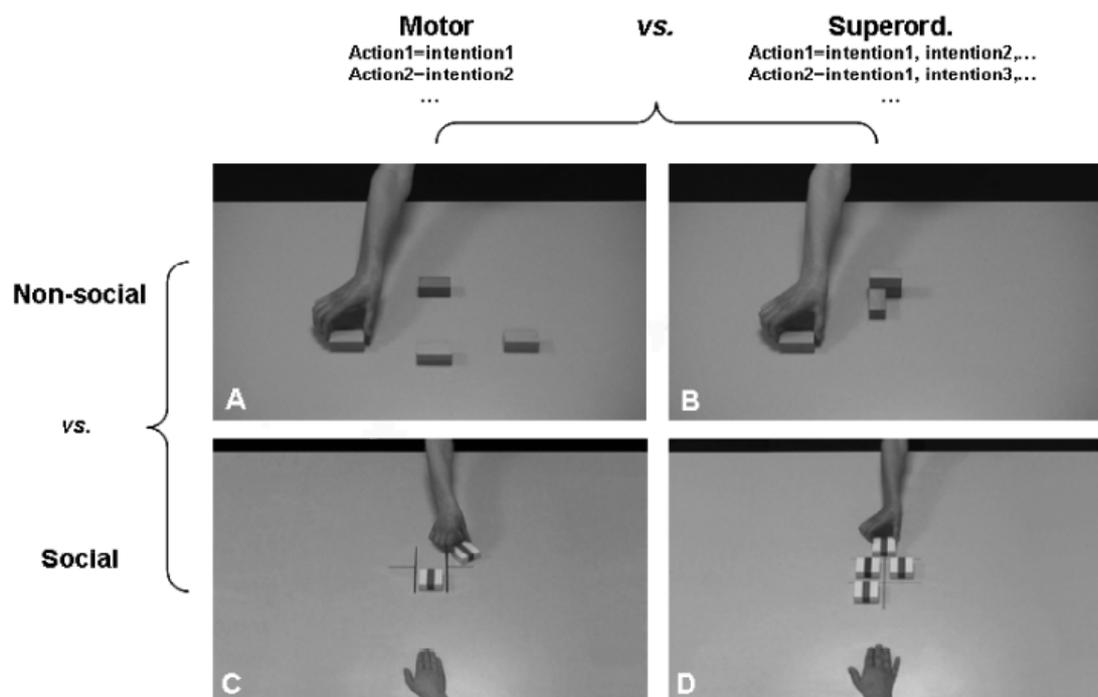


Fig. 2. Examples of stimuli for the motor non-social intention experiment (A); the super-ordinate non-social intention experiment (B); the motor social intention experiment (C) and the superordinate social intention experiment (D).

B. Superordinate experiment: Each trial started with the onset of an incomplete movie showing an actor performing a series of three successive actions on the three cubes available to him. Each individual sequence of actions (*i.e.* each trial) led to building of a final shape composed of those three cubes, and situated in the centre of the scene. Participants were asked to identify the action performed on the last cube, *given* the final shape being currently arranged. Commutative sequences were used so that each shape could be constructed from distinct sequences of actions (e.g. the shape S1 could be obtained from the sequences 'lift-lift-rotate' and 'lift-rotate-lift'). Identifying the last action of the sequence therefore required having previously inferred which superordinate intention (*i.e.* which final shape) this action contributes to achieve. As previously, participants were instructed to indicate what the not-yet completed last action of the sequence was, by pressing as quickly and as accurately as possible, one of the three corresponding keyboard buttons (T, L, or R).

In the bias session, the probability of a particular shape (S1, S2, or S3) was increased at the expense of the two others, while keeping equal the probability of each basic action contributing to building of that shape. The type of the biased shape was counterbalanced across participants.

2.3 Social experiments

The third (figure 2, **C**) and fourth experiments (figure 2, **D**) presented two actors engaged in a social game in which they could either cooperate or defect. Participants were instructed to infer the nature of the second player's social intention (*i.e.* cooperative or defective intention), *given* the action performed by the first player in the previous round (*i.e.* cooperative or defective action). In both these experiments, the bias was assigned according to the way the second player responded to the strategy adopted by the opponent in the previous round. We chose to bias participants toward a 'tit-for-tat' (TFT) mode of reciprocity, rather than toward a particular type of social – defective or cooperative – intention (see Chambon et al., xxxx, for details); meaning that the probability that the second player cooperates if the first player had previously cooperated, or defects if the first player had previously defected, was increased at the expense of the alternative strategies (e.g. always defects, always cooperates). Finally, as in the two previous non-social experiments, both *motor intentions* (single cooperative or defective act) and *superordinate intentions* (sequences of motor acts leading to the construction of a cooperative or defective configuration) were considered.

2.3.1 Material

In both the social motor and superordinate experiment, visual stimuli consisted of incomplete movies showing two players' hands (one in front of the other) manipulating cubes printed with a blue or a red line. The cubes were of similar size (6cm X 6cm) and located at equal distance from the starting position of each player's hand. The first player manipulated the cube closest to him, and then the second player in turn manipulated the cube closest to him. Movies were cut so that the last action (*i.e.* the action performed by the second player) was incomplete (1480 ms, 1560 ms, 1640 ms after the last action starts).

2.3.2. Procedure

C. Social motor experiment: Two cubes (printed with a blue or a red line) were disposed on the sides of a grid that was situated in the centre of the scene. The two actors played one after the other by moving the cube towards the middle of the grid (termed 'bank') or by rotating it so that it remains in its previous place.

Here, each motor act directly denoted the social intention of the player: each player could either cooperate with the other one, by moving the cube towards the central bank (transport), or defect, by leaving the cube in its previous place (rotate). Consequently, there were four possible combinations of intentions, or strategies: either both players cooperated (transport/transport) or defected (rotate/rotate), or the first player defected and the second cooperated (rotate/transport), or the first player cooperated and the second defected (transport/rotate).

For each trial, participants were instructed to observe the incomplete movie and infer what the second player's action was. Participants responded by pressing, as quickly and as accurately as possible, one of the two keyboard buttons corresponding to the two possible actions (T for 'transport', or R for 'rotate') susceptible to achieve the second player's intention (defect or cooperate). Once a response was given, the next trial started.

D. Social superordinate experiment: Four cubes (one blue and three red ones) were placed in the centre of the scene. Actors played in turn with the goal to vertically align three cubes among the four available ones. The goal of the first player was to align the three cubes according to the color red, irrespective of the orientation (configuration 1 or *lco1*). The goal of the second player was to align the cubes according to the orientation, irrespective of the color (configuration 2 or *lco2*). Finally, a third configuration could be obtained by aligning the cubes according to both the orientation and the color (configuration 3 or *lco3*).

To win the game, each player depended upon the other player's manipulation. As in the social motor experiment, the two social intentions were of a defective or a cooperative nature. However, in the present experiment, the social intention was denoted by the sequence of the players' motor acts (i.e. the final configuration: *lco1*, 2 or 3), rather than by the single action performed by each player. Each configuration was associated with a defective or a cooperative strategy that could be equally achieved by rotating or transporting the cube. Each player could therefore adopt a defective strategy by manipulating the cube in such a way that it prevented the creation of a configuration, or adopt a cooperative strategy in order to achieve a configuration. The resulting four possible combinations of strategies therefore led to different final configurations: either both players defected or cooperated (*lco3*), or the first player defected and the second cooperated (*lco1*), or the first player cooperated and the second defected (*lco2*).

Participants were instructed to infer the social superordinate intention of the second player by pressing, as quickly and as accurately as possible, one of the two keyboard buttons corresponding to the two possible actions (T for 'transport', or R for 'rotate') congruent with the second player's social intention (defect or cooperate). Once a response was given, the next trial started.

3. Design and statistical analyses

Three types of analyses were performed between both groups on different dependent variables: 1) Hits (percentage of intentions correctly inferred) and Reaction Times (RTs) for the first group of analyses, 2) a score reflecting the effect of the bias on participants' performances for the second group of analyses, and last, 3) a score accounting for the effect of varying the amount of visual information on participants' performance. Regression analyses were also conducted with the SANS, SAPS and disorganization scores as dependent variables and cognitive performances as explanatory factors.

Finally, for each of the 4 experiments and for each group of participants, two-tailed t-tests were performed between the two non-preferred intentions on both RTs and hits. As no significant differences appeared (all $p > .05$), performances for these two non-preferred intentions were pooled across for subsequent analyses.

3.1. Hits and Reaction Times

In the overt blocks, one two-tailed t-test compared participants' RTs between the first, the second, and the third part of each (baseline and bias) session for non-preferred and preferred intentions. This analysis was conducted to assess the influence of the bias over time.

In the covert blocks, a 2x2x3 repeated-measures ANOVA was performed between groups for each (baseline and bias) session on both RTs and hits. Analyses were made with group (comparison vs. schizophrenic patients) as a between-subjects factor, Bias (preferred vs. non-preferred intention) and Amount of Visual Information (low, moderate and high) as within-subjects factors. Post-hoc Fisher tests were then performed to identify differences between conditions. These same analyses were conducted for each of the 4 experiments. Finally, whenever the variance structure did not conform to the requirements for parametric analyses, logarithmic transformations were used to obtain the required conformity.

For all the analyses, a $p < .05$ was taken as the criterion for significance. These analyses were performed using the statistical software *Statistica 7* (www.statsoft.com).

3.2. Bias effect

To assess the influence of the participants' prior expectations on their performance, and to determine whether this influence varied depending on the type of the intention considered, a 'bias effect' score for both groups was computed across all experiments. This score was obtained by subtracting, in the bias session, the rate of correct responses for the preferred intention to that one of non-preferred intentions. Thus, the greater this score, the more participants responded toward the preferred (*i.e.* biased) intention; that is, the greater they relied on their priors to make their decision.

This score was then input as a dependent variable in a 3x2x4 repeated-measures ANOVA with Amount of Visual Information (low, moderate and high) as a within-subjects factor, and group (comparison vs. schizophrenic patients) and Type of Intention (motor vs. superordinate vs. social motor vs. social superord.) as between-subjects factors.

3.3. Effect of the Amount of Visual Information

Likewise, we calculated a score reflecting the influence of varying the amounts of visual information on participants' performance across all experiments. This score was obtained by subtracting the rate of correct responses obtained in conditions of high amounts to that obtained in conditions of low amounts of visual information, for both preferred and non-preferred intentions. Thus, the greater this score, the more participants' performance was sensitive to increasing the amounts of visual information.

We then input this score in a 2x2x4 repeated-measures ANOVA with Bias (preferred vs. non-preferred intentions) as a within-subjects factor, and group (comparison vs. schizophrenic patients) and Type of Intention (motor vs. superordinate vs. social motor vs. social superord.) as between-subjects factors.

3.4. Regression analyses

Regression analyses were conducted to evaluate the influence of patients' cognitive performance on their clinical scores. In particular, we sought at determining whether an abnormal dependence to the bias and/or to visual information was predictive of the severity of some symptoms dimensions. Independent analyses were thus made with the disorganization, SANS, and SAPS scores as dependent variables, and the 'bias effect' and the 'Visual Information effect' as explanatory factors. For each clinical score, we conducted regression analyses using the different cognitive scores independently (simple linear regressions) or their transformed values (simple non-linear regressions with logarithmic, polynomial or exponential transformations). Models with the highest adjusted R squared (R^2) and a p-value $\leq .05$ are reported.

3. Results

Results of the comparison participants are those obtained in the original study from Chambon et al. (xxxx).

4.1. Hits and Reaction Times (covert blocks)

4.1.1. Overt blocks: constant and very high amount of visual information

All experiments: both groups performed well on the task when the amount of visual information was very high in both the baseline and bias sessions (all mean correct responses percentage > 96%, S.D.<3.4 in healthy participants, and mean > 94%, S.D.<4.1 in schizophrenic patients). These high percentage of correct responses revealed that, as expected, participants did easily recognize the different intentions represented, and by consequence, did integrate the effective probability distribution of each type of (preferred vs. non-preferred) intentions within each (baseline and bias) session.

The figure 3 illustrates the temporal course of patients' RTs over both the sessions: as predicted, the bias was found in the bias session to have a cumulative effect over time, with RTs for preferred intentions progressively decreasing from the first to the last third of the session in all experiments (all $p < .05$).

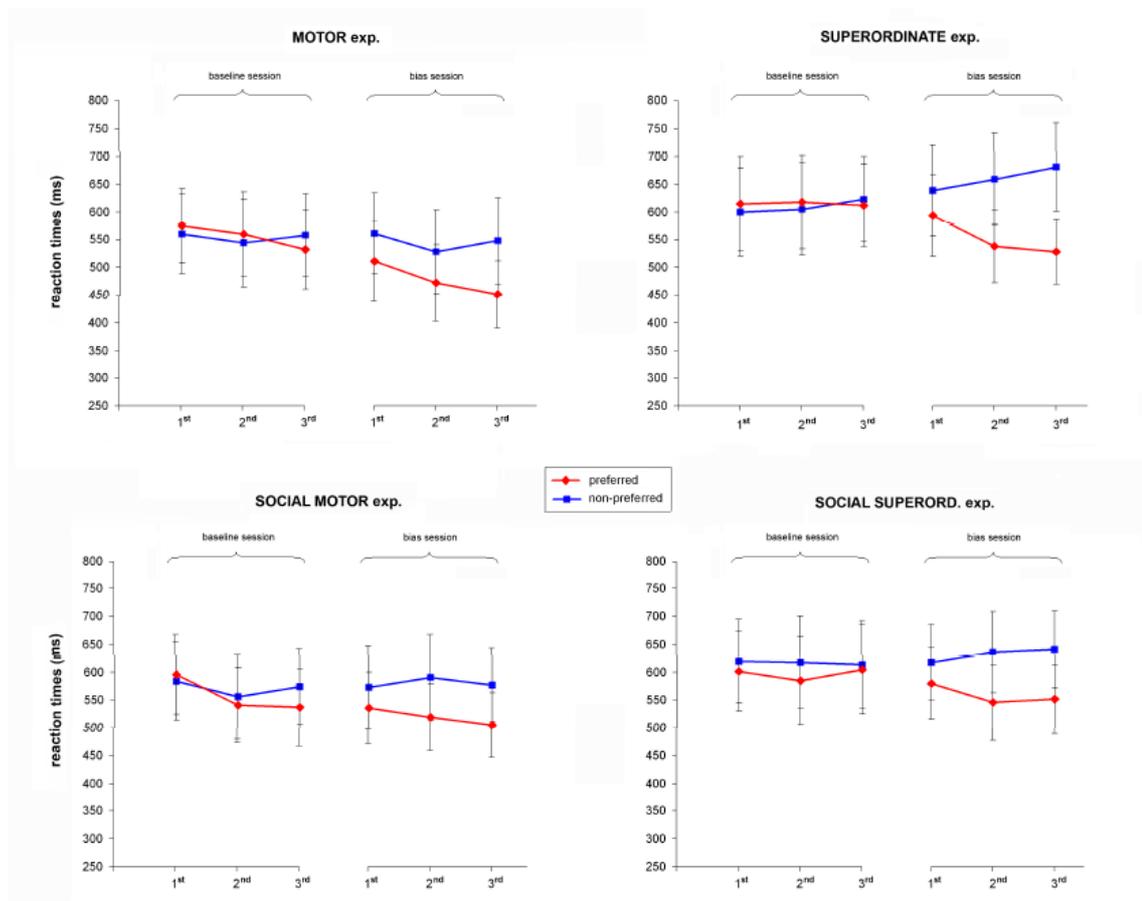


Fig. 3. Schizophrenic patients: Mean reaction times (\pm SD) for the non-preferred and the preferred intentions across time in the OVERT blocks (baseline and bias sessions). 1st: the first three overt blocks of the session; 2nd: the three intermediate blocks; and 3rd: the three last blocks of the session.

4.1.2. Covert blocks: lower and varying amounts of visual information

4.1.2.1. Baseline session (equal probabilities)

Non-social intentions: As previously shown in Chambon et al. (xxxx), healthy participants' RTs decreased and hits improved as the amount of visual information increased. A similar effect was found in schizophrenic patients (all participants, both motor and Superord. tasks: $F(2,96) > 251.1$, $p < .001$). Moreover, both groups showed no significant differences in correct responses between all types of intentions presented, indicating that prior to biasing participants, there wasn't any *a priori* preference for one intention over the other ones (both tasks: $F(1,48) > .01$, $p > .85$). Finally, we didn't find any interaction between the amount of

visual information and the bias in both comparison and schizophrenic participants, showing that improved participants' performance for higher amounts of information was independent of the type of intention presented (both tasks: $F(2,96) > 2.3$, $p > .12$).

Social intentions: As for non-social intentions, a significant main effect of the amount of visual information was obtained on participants' hits and RTs, in both comparison and schizophrenic participants (both social motor and social superord. tasks: $F(2,94-96) > 159.9$, $p < .001$). However, we found groups' differences in the main effect of the factor 'Bias': while healthy participants showed an early preference (*i.e.* prior to being biased) for 'tit-for-tat' mode of reciprocity, patients did not. In the social superordinate task, a significant effect of the interaction Group*Bias was thus obtained for hits rate ($F(1,48) = 6.3$, $p = .01$), with patients showing no significant preference for intentions congruent with a 'tit-for-tat' strategy compared to the other ones (*e.g.* 'always cooperate', 'always defect' strategies). In the social motor task, this interaction did not reach significance however ($p = .09$).

4.1.2.2. Bias session (biased probabilities)

Non-social intentions: In both groups, a significant main effect of the amount of visual information was obtained (both tasks: $F(2,96) > 192$, $p < .001$), as well as a main effect of the Bias (both tasks: $F(1,48) > 133.8$, $p < .001$) indicating that healthy subjects as schizophrenic patients preferentially responded toward intentions with accrued probability. In the motor experiment, we did not find any interaction between the group and Bias factors, with patients performing as comparison participants ($F(1,48) = 0.42$, $p = .51$), nor between group and Amount of visual information ($F(2,96) = 0.01$, $p = .9$). In the superordinate task, however, a significant interaction effect was observed between the group and Bias factors on one hand ($F(1,48) = 5.4$, $p = .02$), and between groups and the Amount of visual information on the other hand ($F(2,96) = 3.2$, $p = .04$). In this condition, patients' performances improved to a lesser extent than those of healthy when the amount of information increased; furthermore, patients were more sensitive to the bias than comparison participants. Fisher tests further revealed that this increased dependence on the bias was accounted for by the fact that, while patients and healthy performed at a comparable level for recognizing preferred intentions, patients provided less responses toward non-preferred intentions than controls participants (LSD test: $p < .001$).

Social intentions: In both the social motor and superordinate tasks, we also observed a main effect of the amount of visual information, together with a significant effect of the Bias: both groups' performances were found to improve as the amount of information increased (both tasks: $F(2,94-96) > 181.8$, $p < .001$), and participants provided a greater number of correct responses for intentions congruent with the preferred mode of reciprocity (TFT) than with non-preferred ones ($F(1,47-48) > 33.4$, $p < .001$). We also observed a significant effect of interaction between groups and the amount of visual information: to the contrary of non-social experiments, patients' performances improved to a greater extent than those of healthy when the amount of information increased (both tasks: $F(2,94-96) > 3.11$, $p < .05$). In the social superordinate experiment, the effect of interaction between group and Bias was also significant, with patients less relying on their prior expectations than healthy participants ($F(1,48) = 6.7$, $p = .01$). This result was accounted by the fact that patients provided less correct responses toward preferred intentions than control participants (LSD test: $p = .04$), while performing at a comparable level in inferring non-preferred ones (LSD test: $p = .62$).

4.2. Bias effect

There was no main effect of the group, indicating that, regardless to the type of intention, the effect of the bias was as important in patients as in healthy participants ($F(1,191)=2.1$, $p=.15$). The main effect of the Amount of visual information was significant, with an effect of the bias progressively increasing as the amount of information decreased ($F(2,382)=93.8$, $p<.001$). We also observed a main effect of the Type of intention ($F(3,191)=11.7$, $p<.001$), accounting for a greater influence of the bias for inferring superordinate vs. motor intentions, whatever the type of target considered (social as non-social). Finally, the effect of interaction between the Type of intention and groups was significant ($F(3,191)=11.06$, $p<.001$). Interestingly, patients' performance testified to a greater influence of the bias than controls in the non-social superordinate condition (LSD test: $p=.003$), while in social conditions the bias affected patients' response to a lesser extent than in healthy participants (LSD tests: all $p<.008$) (figure 4). This interaction effect was not further modulated by the Amount of visual information (Group*Amount*Intention: $F(6,382)=.69$, $p=.66$).

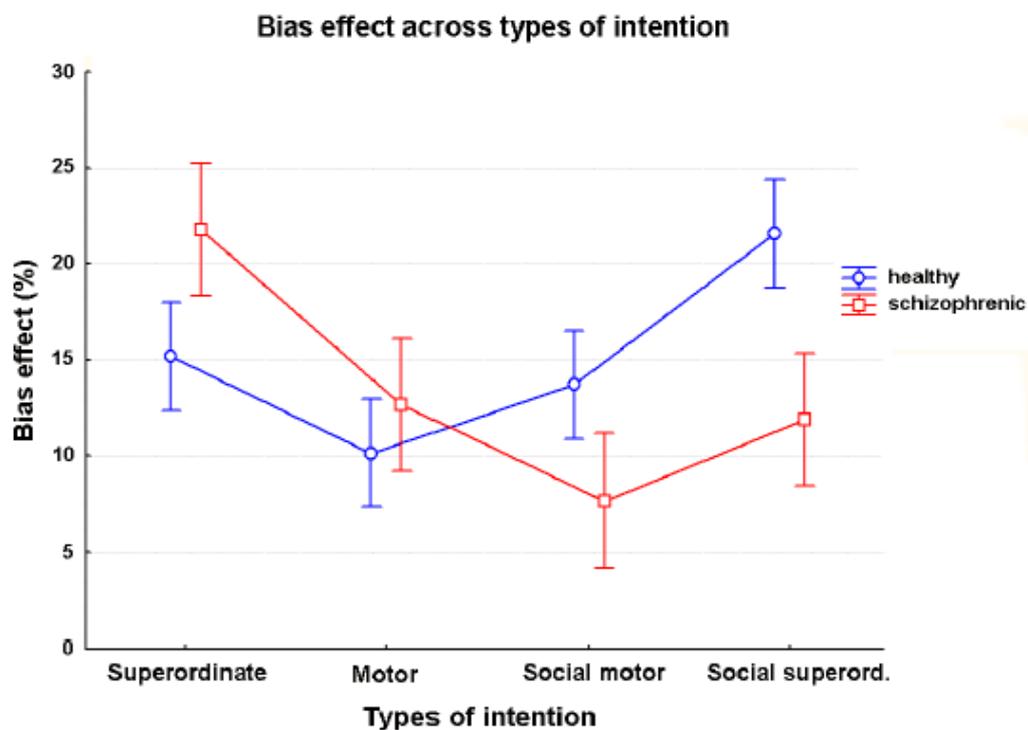


Fig. 4. Bias effect (%) for all types of intention considered. The greater the bias effect, the more participants respond toward the preferred (i.e. biased) intention.

4.3. Effect of the Amount of Visual Information

The main effect of group was significant, with patients' performance improving to a greater extent than healthy participants' when increasing the amount of visual information ($F(1,191)=4.12$, $p=.04$). We did not find any significant main effect of the Type of intention ($F(3,191)=2.6$, $p>.05$). However, a significant interaction effect between group and Type of intention was obtained ($F(3,191)=6.12$, $p<.001$): in non-social conditions, patients' response was equally (motor), if not less (superordinate), affected than controls' one by varying the amounts of visual information (LSD test: $p=.03$), while in social conditions, they relied significantly more on these amounts than healthy (LSD tests, both tasks: all $p<.005$) (figure 5).

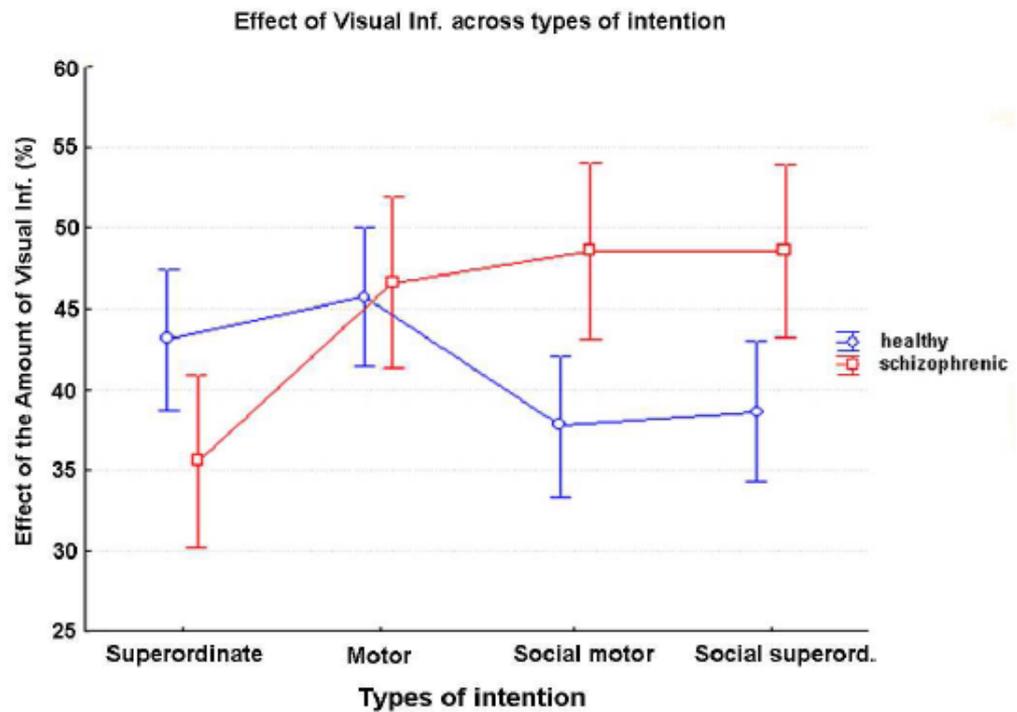


Fig. 5. Effect of the Amount of Visual Information (%) for all types of intention considered. The greater this score, the more participants' performance are sensitive to increasing the amounts of visual information.

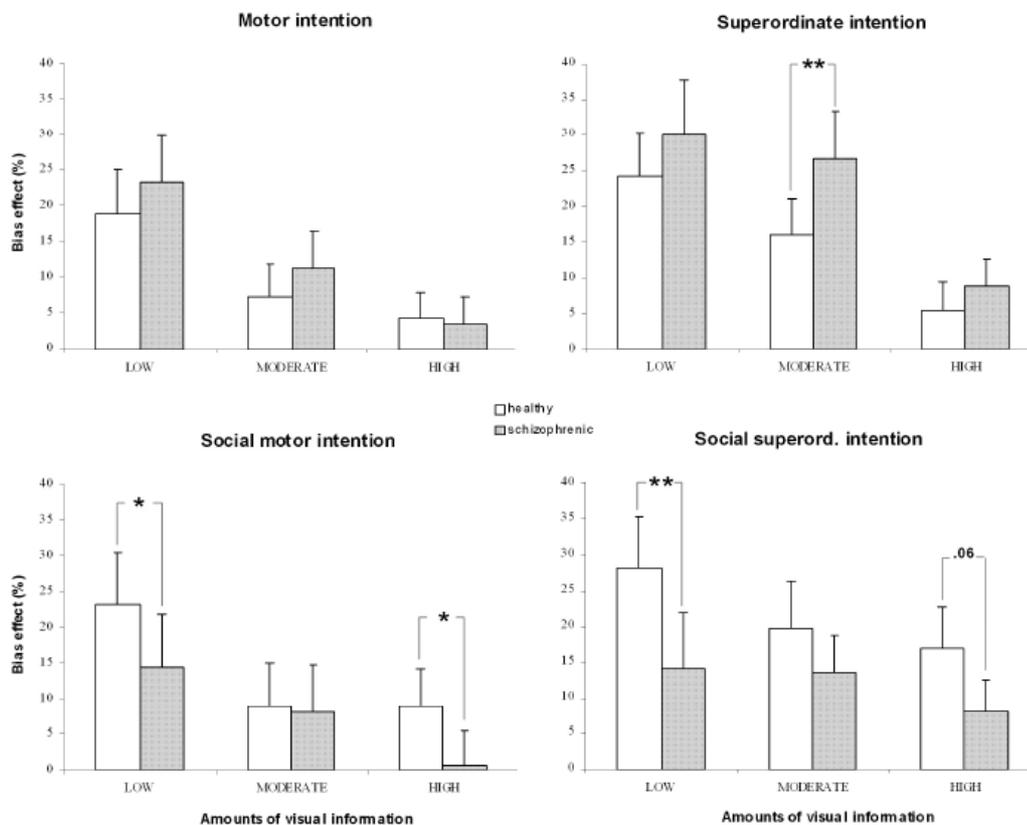


Fig. 6. Both groups: Bias effect for the 3 amounts of visual information considered (low, moderate, high), across all types of intention. Between-group differences marked with * = $p < .05$, or ** = $p < .01$.

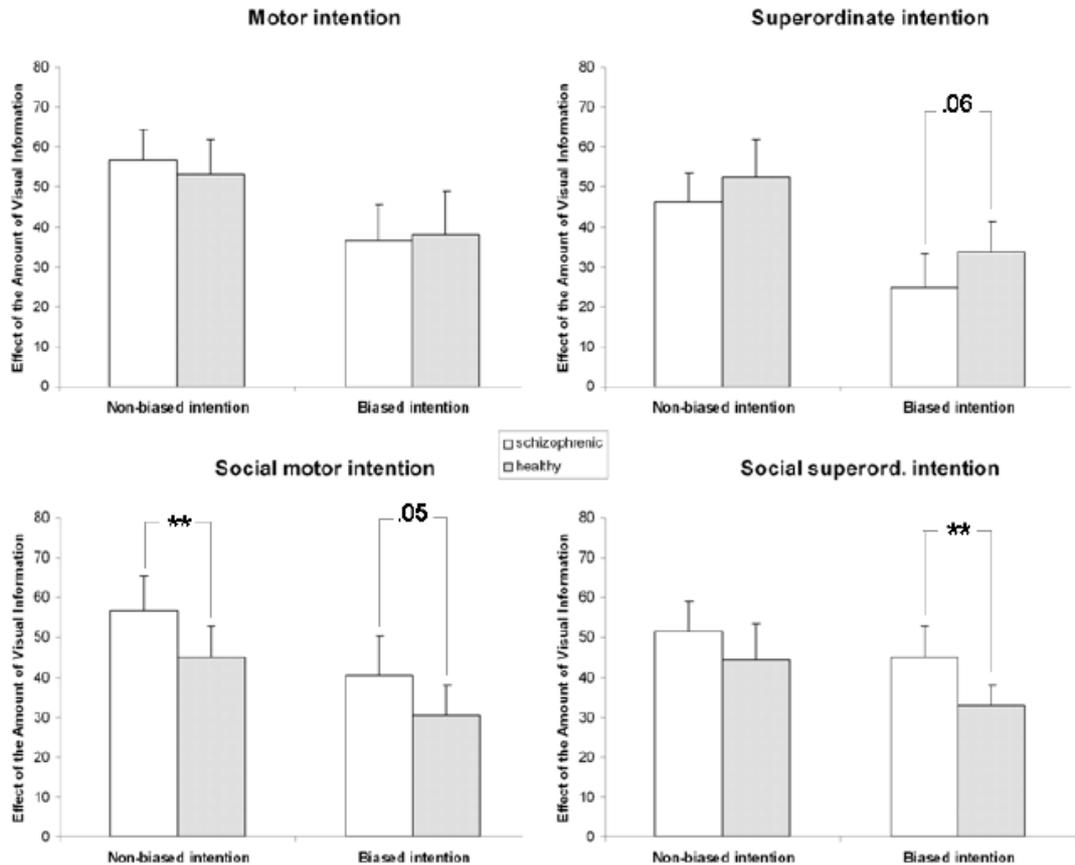


Fig. 7. Both groups: effect of the Amount of Visual Information on inferring non-biased and biased intentions, across all conditions. Between-group differences marked with * = $p < .05$, or ** = $p < .01$.

4.3 Regression analyses

4.3.1 Bias effect and clinical scores

In the non-social motor task, neither the SANS, the SAPS, or the disorganization scores were found to be predicted by the Bias effect. In the non-social superordinate task, the Bias effect significantly predicted the disorganization score ($R^2=0.21$, $p=.04$), and the SAPS score ($R^2=0.39$, $p=.003$) (figure 8, a. and b.). The more patients were disorganized and exhibited *positive* symptoms, the higher was the Bias effect on their performances, *i.e.* the more they relied on their priors to make their decision.

In the social motor task, the Bias effect was found to significantly predict the SANS score ($R^2=0.22, p=.03$), while in the social superordinate task, the Bias effect also predicted the SANS score ($R^2=0.32, p=.008$) (figure 8, c. and d.) and tended to explain a part of the variance in the disorganization score ($R^2=0.18, p=.09$). Thus, the more patients exhibited *negative* symptoms, the smaller was the Bias effect on their performances, *i.e.* the less they relied on their priors when inferring intentions accomplished in a context of social interaction.

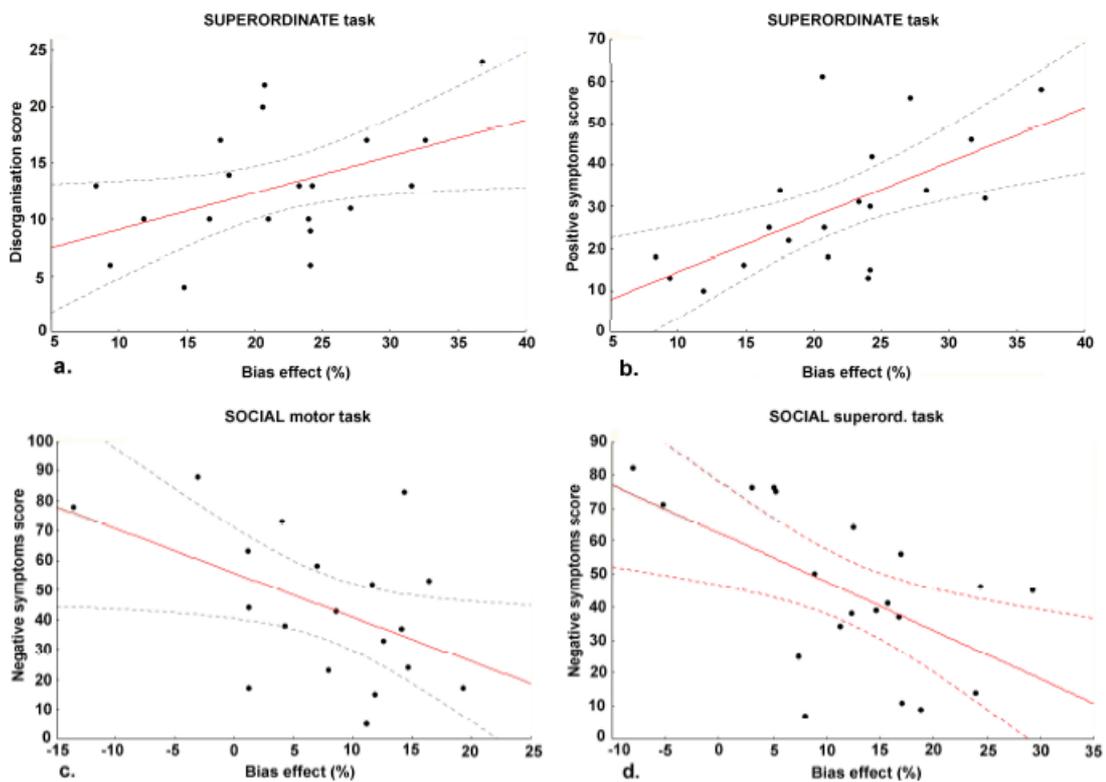


Fig. 8. The linear regression lines derived from the linear regressions analyses between the Bias effect (explanatory factor) and patients' disorganization score (a), positive (b) and negative symptoms (c,d) are shown in red. The 95% confidence intervals around the regression lines are shown in grey.

4.3.2. Amount of visual information and clinical scores

In both the non-social tasks, no symptoms dimension was found to be predicted by the effect 'Amount of Visual Information'.

In the social motor task, this effect only tends to predict the SANS score ($R^2=0.17$, $p=.07$) while in the social superordinate condition, the Amount of Visual Information effect significantly predicted both the SANS ($R^2=0.44$, $p=.001$) and the disorganization ($R^2=0.28$, $p=.01$) scores (figure 9). The more patients were disorganized and the more severe were their negative symptoms, the higher was the effect of Visual Information on their response, *i.e.* the more they relied on the visual information to make their decision.

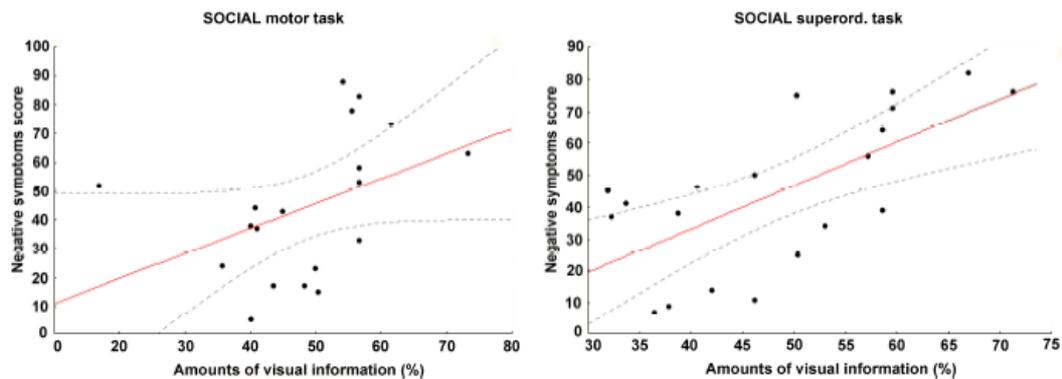


Fig. 9. Social tasks: The linear regression lines derived from the linear regressions analyses between the effect of the Amount of Visual Information (explanatory factor) and patients' negative symptoms are shown in red. The 95% confidence intervals around the regression lines are shown in grey.

5. Discussion

The present study aimed to investigate whether schizophrenic patients' impaired ability in appreciating others' intentions is confined, rather than generalized, to a particular type of intention. To test this hypothesis, we designed a series of tasks requiring to identify different types of intentions, varying according to either their scope or target. Our second hypothesis was that a deficit in intention attribution, if established, should be accounted for by an abnormal interplay of the two types of information which usually contribute to this ability. In a previous study, we indeed showed that appreciating others' intentions relied on an interaction between visual information – derived from the agent's kinematics – and prior knowledge that participants progressively gained about the most likely causes of what they observed (Chambon et al., xxxx).

The results revealed that patients' performance accounts for this interaction whatever the type of intention considered. First, patients better recognized the underlying intentions when the visual information conveyed by the action scene increased; second, they provided a greater number of correct responses and faster RTs for preferred intentions than for non-preferred ones; finally, in both groups, performances exhibited a strong bias effect which progressively increased as the amount of visual information decreased, and vice versa.

5.1. Inferring non-social intentions

In terms of performance, overall, the only between-group difference was observed in the superordinate condition, in which participants had to appreciate distal goals of an agent performing a sequence of basic motor acts. In that condition, patients with schizophrenia were found to perform significantly less well than healthy participants. Importantly, these

difficulties were associated with an abnormally high effect of the bias, with patients providing significantly less responses than controls toward the non-preferred intentions than toward the preferred ones. Last, in the schizophrenia group, performances improved to a lesser extent than in healthy controls as the visual information increased, further suggesting that patients' tendency to over-rely on their internal, prior expectations arise together with a disposition to neglect the external, visual information conveyed by the action scene.

This pattern of performance might be the consequence of an impaired ability to revise expectations which one progressively formed during the task. Increasing the probability of some intentions indeed led participants to favour specific expectations about which type of intention was the most likely to be accomplished, given past observation. Because preferred and non-preferred intentions were randomly alternated in the sequences, participants had to continuously revise their expectations regarding the type of intention to come. The patients' inclination to persevere in selecting preferred intentions may suggest an impaired ability to revise these expectations, *i.e.* to update their beliefs about the agent's current intention. Concomitantly, schizophrenic patients showed an information-gathering bias that led them to mistrust the visual information which could potentially disconfirm these initial beliefs.

Such an inability to disengage from prior, self-generated expectations, together with a tendency to ignore outer information, echo specific biases observed in schizophrenia in a wide range of studies, such as a 'confirmatory bias' (Jones et al., 1999), or a tendency to make hasty decisions (Colbert et al., 2002; Huq et al., 1988; Brankovic & Paunovic, 1999). In tasks of probabilistic reasoning, patients with delusions indeed tend to make a decision on the basis of less evidence than healthy controls. They are also overconfident in their own judgments relative to healthy participants (Huq et al., 1988) and preferentially seek evidence that confirms their initial beliefs (Brankovic & Paunovic, 1999). More generally, deluded patients are prone to give excessive credit to their own internal productions, and, as a consequence, to hold them despite little evidential support (Jones et al., 1999).

These observations fit particularly well with the present results: in the superordinate condition, schizophrenic patients are prone to give massive priority to biased intentions at the expense of non-biased ones and to underweight visual evidence arising from the action scene. Consistently, we found this profile of performance to significantly predict the severity of both the productive and disorganisation symptoms of the disease. Previous evidence showed that productive patients may be characterised by an impaired ability in decoupling their own intentions from those of others (Frith & Corcoran, 1996), while disorganised individuals exhibit specific difficulties in organizing their actions in the context of their own, but also other people's goals (Zalla et al. 2006 ; Chambon et al., 2008). The present results further suggest that productive, as well as disorganised patients, exhibit a mentalizing profile that may be characterized by a disposition to hold prior beliefs about others' intention with an abnormal degree of certainty, making them incorrigible by external evidence; thereby outdated beliefs about others' mental states are not replaced by contextually appropriate ones. Finally, we show that such a profile (over-reliance on priors, disconfirming visual evidence) is characterised by specific difficulties in inferring distal, rather than proximal intentions. The reason for this specificity may be that inferring superordinate intentions requires participants to refer to an internal representation of the goal which is not directly available from observation, and, as a result, tends to be less challenged by visual evidence.

In total, depending on outdated expectations could characterise an abnormal style of reasoning that would contribute in both reasoning biases observed in deluded patients and specific mentalizing disorders shown in productive patients in general. As such, failure to revise prior expectations about other people's goals and intentions could further

account for productive patients misattributing actions, especially in situations that require continuously monitoring visual signals arising from the action scene (Franck et al., 2001). Finally, patients' difficulties in disengaging from their prior expectations might echo problems that schizophrenics with passivity symptoms have in decoupling outer from inner world experience (Blakemore, 2003; Frith, Farrer et Franck, etc.), resulting in an abnormal inclination to attribute intentionality where there is none ('over-TOM': Abu-Akel & Bailey, 2000; Blakemore et al., 2003; Franck et al., 2001), or to perceive moving geometrical objects as endowed with agency (Frith, 2005).

5.2. Inferring social intentions

In social tasks, patients overall tended to perform less well than healthy controls, but this tendency did not reach significance. As expected, schizophrenic patients were also found to significantly favour the 'tit-for-tat' (TFT) mode of reciprocity, by preferentially attributing intentions congruent with that mode to the second agent (*i.e.* defect if previous defection, cooperate if previous cooperation) to the expense of others (*i.e.* always defecting, always cooperating). However, two main results made them diverge from healthy participants. First, while healthy participants showed an early preference toward TFT intentions in the baseline session (that is, prior to being biased), patients did not. As a result, patients were found to provide lesser responses toward TFT intentions than controls did in the bias session. Second, this lesser interest for TFT intentions was concomitantly associated with the number of correct responses improving to a greater extent than those of healthy participants as the amount of visual information increased. Finally, the more the negative symptoms were severe, the less patients were sensitive to the bias, *i.e.* the less they responded toward the preferred mode of reciprocity. Accordingly, the more the negative symptoms were severe, the more patients relied on the visual information to make their decision.

Crucially, this pattern of performance was the exact opposite of the one we observed in the non-social tasks. One explanation for this result might reside in the fact that social conditions – which required paying attention to both the first and second player's intention –, were more demanding than non-social ones. However, if that was the case, poorer performances should have been observed in these conditions. Yet, schizophrenic patients performed at comparable levels across non-social and social experiments, indicating that the lesser effect of the bias on patients' social decision is accounted for by the type of intention considered rather than by the difficulty of the task.

Two possible explanations may account for why the bias did not influence patients' social inference to the same extent as it did for that of controls'. In healthy participants, situations identified as involving social interactions are generally prone to trigger domain-specific expectations concerning the way agents are likely to behave in such situations (Frith & Frith, 2006; Castelli, Happé, Frith & Frith, 2000). In the present study, healthy participants exhibited such expectations as early as the baseline session – with faster RTs in predicting intentions congruent with the TFT mode of reciprocity –, while patients did not. Moreover, we found that the more the negative symptoms were severe, the fewer patients responded toward this mode of reciprocity. It is therefore possible that social situations do not induce the same domain-specific expectations in negative patients which are normally observed in healthy participants. Impoverished social knowledge has been consistently argued to be an intrinsic feature of schizophrenia (Cutting & Murphy, 1990). Previous evidence also suggests that observable behavioural information derived from social interaction is abnormally used by negative patients in problem solving tasks (Bedell et al., 1998). The present results suggest, accordingly, that patients with severe negative symptoms do not properly use social

domain-knowledge to predict what interacting agents are most likely to do. This is further echoed by the fact that patients compensated for this by overly relying on visual evidence.

Such over-reliance on visual, outer information, together with impoverished social expectations is consistent with the observation that patients with negative symptoms – such as anhedonia or alexithymia – tend to excessively focus on directly observable, external information, rather than inner experience (Taylor, 1994). However, fewer social expectations in schizophrenia do not necessarily imply that patients have none. Their responses in the bias session reveal that they are able to acquire and use knowledge about social situations. This suggests an alternative explanation for their pattern of performance: in a series of ToM tasks with stories depicting interacting agents, Sarfati and colleagues have shown that patients preferentially based their interpretation of intention on socially familiar experiences rather than on the context of the story (Sarfati et al., 1997). Drury and coworkers also found that schizophrenic patients are prone to be distracted by the relations between the story characters and their own personal experiences (Drury et al., 1998). One cannot therefore preclude that patients do have strong expectations of how agents are supposed to interact, but that these expectations are abnormally related to their own (inappropriate?) social experience. Selecting social intentions congruent with the context of the task would therefore require them to disengage from these prior expectations. This would be a costly process which in our task could be accounted for by patients' increased reaction times toward TFT intentions.

6. Conclusion

Inferring other people's intentions requires properly dealing with both the visual information from the action scene and prior knowledge about the most likely causes of what is observed. Schizophrenic patients exhibited some abnormal dependence on both these types of information. In particular, patients' prior knowledge was found to abnormally modulate their decision about other's action depending on the type of intention considered and their prevalent symptoms.

We first showed that patients with severe positive symptoms overly relied on their priors, but also on disconfirming visual evidence from the action scene. This pattern of response, resulting in overgenerating internal hypotheses and dismissing potentially disconfirmatory evidence, may favour in these patients a paranoid interpretation of other people's goals. It may also result in abnormal distinguishing between one's own and other's intention, a confusion that patients with passivity symptoms frequently experience, such as in the well-documented delusion of control.

In social conditions, on the other hand, schizophrenic patients exhibited an inappropriate use of social domain-knowledge. This biased knowledge was found to induce patients with inappropriate expectations of how agents were the most likely to behave in a context of social interaction. Patients with negative symptoms compensated for this by giving priority to visual evidence in making their decision. We suggested that this profile of performance parallels the stimulus-induced nature of negative behaviours in schizophrenia and could further account for action initiation disorders commonly observed in this dimension of the disease.

Acknowledgments

This research was supported by a grant of the Agence Nationale de la Recherche (ANR-ZR81) and by a grant of the Conseil Scientifique de Recherche, Le Vinatier (CSR U03). V. Chambon was supported by scholarships from the French Ministry of Research. This research was approved by the local Ethical Committee (B80631-60). The authors report no competing interests. We are grateful to Mathilde Vernet, Juliette Sablier and Mariia Kaliuzhna for their help in preparing the stimuli. We also gratefully acknowledge Mrs Ghislaine Bailly, Drs Olivier Martin and Caroline Demily, and Pr. Jean-Louis Terra for their help in the inclusion of patients.

References

- Abu-Akel, A., & Bailey, A.L. (2000). The possibility of different forms of theory of mind, *Psychological Medicine*, 30, 735-738.
- Andreasen, N.C. (1983). The Scale for the Assessment of Negative Symptoms (SANS), Iowa City, The University of Iowa.
- Andreasen NC. (1984). The Scale for the Assessment of Positive Symptoms (SAPS), Iowa City, The University of Iowa.
- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2006). Bayesian Models of Human Action Understanding. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18, pp. 99-106), MIT Press.
- Bedell, J., Lennox, S.S., Smith, A.D., & Rabinowicz, E.F. (1998). Evaluation of problem solving and communication skills of persons with schizophrenia, *Psychiatry Res.*, 78, 197-206.
- Blakemore, S.J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The detection of contingency and animacy from simple animations in the human brain, *Cerebral cortex*, 13, 837-844.
- Brankovic, S.B., & Paunovic, V.R. (1999). Reasoning under uncertainty in deluded schizophrenic patients: A longitudinal study, *European Psychiatry*, 14, 76-83.
- Brunet, E. Sarfati, Y., & Hardy-Bayle, M.C. (2003). Reasoning about physical causality and other's intentions in schizophrenia, *Cogn Neuropsychiatry*, 8, 129-139.
- Castelli, F., Happé, F., Frith, U. & Frith, C. D. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns, *Neuroimage*, 12, 314-325.
- Chambon V., Franck N., Koechlin E., Ciuperca G., Fakra E., Azorin J.-M., & Farrer C. (2008). The architecture of cognitive control in schizophrenia, *Brain*, 131, 962-970.
- Chambon, V., Domenech, P., Pacherie, E., Koechlin, E., Baraduc, P., & Farrer, C. (xxxx). What are they up to? The role of sensory evidence and prior knowledge in action understanding.
- Colbert, S.M., & Peters, E.R. (2002). Need for closure and jumping-to-conclusions in delusion-prone individuals, *Journal of Nervous and Mental Disease*, 190, 27-31.
- Corcoran, R. (2001). Theory of Mind in Schizophrenia, In Penn, D., Corrigan, P. (eds), *Social Cognition in Schizophrenia*, APA, Washington.
- Cutting, J. & Murphy, D. (1990). Impaired ability of schizophrenic, relative to manics or depressives, to appreciate social knowledge about their culture, *Br. J. Psychiatry*, 157, 355-358.

- Drury, V.M., Robinson, E.J., & Birchwood, M. (1998). 'Theory of mind' skills during an acute episode of psychosis and following recovery, *Psychological Medicine*, 28, 1101-1112.
- Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia, *American Journal of Psychiatry*, 158, 454-459.
- Frith, C.D. (1992). *The cognitive neuropsychology of schizophrenia*, Hove, Lawrence Erlbaum Associates.
- Frith, C.D., & Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia, *Psychological Medicine*, 26, 521-530.
- Frith, C.D. (2004). Schizophrenia and theory of mind, *Psychological Medicine*, 34, 385-389.
- Frith, C.D. (2005). The self in action: Lessons from delusions of control, *Conscious Cognitive*, 14, 752-770.
- Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do, *Brain Research*, 1079, 36-46.
- Harrington, L., Siegert, R. J. & McClure, J. (2005) Theory of mind in schizophrenia: a critical review. *Cognitive Neuropsychiatry*, 10, 249 -286.
- Huq, S.F., Garety, P.A., & Hemsley, D.R. (1988). Probabilistic judgements in deluded and non-deluded subjects, *Quarterly Journal of Experimental Psychology: Human Learning and Memory*, 40, 801-812.
- Jones, E. (1999). The phenomenology of abnormal belief. *Philosophy, Psychiatry and Psychology*, 6, 1-16.
- Langdon, R., Coltheart, M., Ward, P.B., & Catts, S.V. (2002). Disturbed communication in schizophrenia: The role of poor pragmatics and poor mind-reading, *Psychological Medicine*, 32, 1273-1284.
- Mazza, M., De Risio, A., Surian, L., Roncone, R., & Casacchia, M. (2001). Selective impairment of theory of mind in people with schizophrenia, *Schizophrenia Research*, 47, 299-308.
- McCabe, R. (2004). On the Inadequacies of Theory of Mind Explanations of Schizophrenia. *Alternative Accounts of Alternative Problems, Theory & Psychology*, 14, 738-752.
- Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory, *Neuropsychologia*, 9, 97-113.
- Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychol*, 124, 106-28.
- Pacherie, E. (2000). The content of intentions. *Mind and Language*, 15, 400-432.
- Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition*, 107, 179-217.
- Pickup, G.J., & Frith, C.D. (2001). Theory of mind impairments in schizophrenia: symptomatology, severity and specificity, *Psychological Medicine*, 31, 207-220.
- Sarfati, Y., Hardy-Bayle, M.C., Nadel, J., Chevalier, J.-FR., & Widlocher, D. (1997). Attribution of mental states to others by schizophrenic patients, *Cognitive Neuropsychiatry*, 2, 1-17.

- Sarfati, Y., Hardy-Bayle, M.C., Brunet, E., & Widlocher, D. (1999). Investigating theory of mind in schizophrenia: influence of verbalization in disorganized and non-disorganized patients, *Schizophrenia Research*, 25, 183-190.
- Sprong, M., Schothorst, P., Vost, E., Hox, J., & van Engeland, H. (2007). Theory of mind in schizophrenia: a meta-analysis, *British Journal of Psychiatry*, 191, 5-13.
- Taylor, G.J. (1994). The alexithymia construct: conceptualization, validation, and relationship with basic dimensions of personality, *New Trends Exp Clin Psychiatry*, 10, 61-74.
- Walter, H., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R.B., Erk, S., & Bara, B.G. (2009). Dysfunction of the social brain in schizophrenia is modulated by intention type: An fMRI type, *SCAN*, 4, 166-176.
- Zalla, T., Verlut, I., Franck, N., Puzenat, D., & Sirigu, A. (2004). Perception of dynamic action in patients with schizophrenia, *Psychiatry Research*, 128,39-51.
- Zalla, T., Bouchilloux, N., Labruyere, N., Georgieff, N., Bougerol, T., & Franck, N. (2006). Impairment in event sequencing in disorganised and non-disorganised patients with schizophrenia. *Brain Research Bulletin*, 68,195-202.

5. Discussion générale

5.1 Le rôle de l'information perceptive et des attentes *a priori* dans la compréhension de l'action

Les résultats obtenus chez les sujets sains sont conformes aux prédictions que nous avons formulées. Dans les quatre études, la quantité d'information visuelle a un effet significatif sur les performances des participants : lorsque la durée des séquences d'action augmente, l'intention sous-jacente est naturellement plus facile à identifier. Les participants sont également sensibles à la structure probabiliste des séquences et tendent à répondre préférentiellement en direction des intentions dont la probabilité d'apparition est, sur l'ensemble de la séquence, la plus élevée. Cette augmentation des performances en direction de l'intention « privilégiée » interagit par ailleurs avec la quantité d'information visuelle disponible : dans la condition *motrice* (lever, tirer ou tourner le cube), et lorsque les participants disposent de peu d'information visuelle sur l'action, ils tendent à répondre préférentiellement en direction de leurs *a priori* – l'intention dont la probabilité est la plus élevée. Dans les conditions où l'intention à reconnaître est de nature « privée » (*superordinate*), ce phénomène de dépendance est amplifié et vient progressivement contaminer les niveaux intermédiaires d'information visuelle : les participants continuent de se référer à leurs *a priori* (l'intention qu'ils estiment *a priori* la plus probable) tout en négligeant les informations perceptives à leur disposition. Enfin, l'ajout d'une composante sociale témoigne d'une amplification de ce phénomène, avec une tendance marquée à répondre en direction des stratégies biaisées, même dans les conditions où l'information visuelle disponible est élevée et peut, à l'occasion, contredire les attentes préalables des participants.

Les résultats de cette première série d'études fournissent au total plusieurs indications. Reconnaître une intention, d'une part, mobilise des processus dont le fonctionnement

s'ajuste au cadre plus général du formalisme Bayésien. Nous avons en effet montré que l'inférence intentionnelle pouvait être modélisée comme le produit d'une interaction entre des informations *a priori* – qui ont une certaine probabilité – et des informations perceptives – qui ont, elles, une certaine vraisemblance, variable selon les conditions. Lorsqu'on réduit progressivement la vraisemblance de ces informations perceptives (en manipulant la durée des séquences d'action), les sujets tendent à répondre de plus en plus massivement en direction des intentions qu'ils estiment les plus probables. Ces résultats confortent l'hypothèse selon laquelle les situations intentionnelles peuvent être interprétées comme des *problèmes inverses*, c'est-à-dire des situations – dans notre cas, des scènes d'actions – dont les causes, potentiellement multiples, ne peuvent être déterminées sur la foi des seules informations perceptives disponibles (Baker et al., 2006 ; Kilner et al., 2007 ; Keysers & Perrett, 2004 ; Wolpert et al., 2003). Or, confronté à un problème de ce type, il a été montré les sujets tendaient à résoudre l'ambiguïté en recourant à des hypothèses *a priori* sur la nature du phénomène observé. Ce recours a été observé en perception pure, pour la résolution de stimuli bi-stables (Mamassian & Goutcher 2001), mais également en situation d'interprétation téléologique. Les enfants, par exemple, manifestent des attentes *a priori* extrêmement robustes sur les causes probables de certains événements, et ces attentes s'expriment conjointement à une tendance irrépressible à interpréter certains événements comme étant intentionnels, c'est-à-dire dirigés vers un but ('teleological obsession'). Lorsque l'information perceptive n'est pas suffisante pour interpréter un événement comme dirigé vers un but (Csibra et al., 1999), ou lorsque l'action n'est que partiellement réalisée (Onishi et al., 2007), les enfants vont jusqu'à postuler des états du monde qui contredisent l'évidence perçue (tels que la présence d'objets physiques cachés). Ces hypothèses viennent compléter, sinon corriger, l'incomplétude perceptive de la situation et confirmer l'hypothèse *a priori* que l'action observée est bien intentionnelle. Nous croyons que les résultats de la présente étude traduisent, pour tous les niveaux intentionnels évalués, la mise en œuvre d'un mécanisme de complétion *a priori* de ce type.

Second résultat d'importance, ce recours massif aux *a priori* s'avère par ailleurs sensible aux variations de complexité intentionnelle : les participants tendent en effet à se référer davantage à leurs *a priori* lorsqu'on progresse dans la typologie – des intentions motrices simples aux intentions superordonnées, puis sociales. Nous avons fait l'hypothèse que cette différence tenait à la nature-même des intentions présentées. Dans les conditions « superordonnées », les séquences d'actions sont commutatives, et, de fait, un même acte moteur (« tourner » par exemple) peut indifféremment réaliser deux intentions distinctes (construire la forme S1, construire la forme S3)⁸. Ici, l'acte observé entretient donc une relation de correspondance multiple avec l'intention qu'il réalise (*one-to-many relation*), au contraire des conditions motrices où l'acte simple (imprimer une rotation au cube par exemple) dénote directement, et sans ambiguïté, l'intention sous-jacente (« tourner le cube ») (*one-to-one relation*). En conditions superordonnées, l'information visuelle véhiculée par la kinématique de l'acte moteur n'est donc pas suffisante pour prédire sans ambiguïté l'intention superordonnée que cet acte réalise. L'information perceptive est en ce cas naturellement ambiguë : elle sous-détermine l'ensemble des buts possibles qu'on peut raisonnablement lui associer. Or, en contexte d'incertitude visuelle, les participants n'ont pas d'autres choix que de se référer à leurs connaissances préalables, c'est-à-dire privilégier l'intention superordonnée qu'ils estiment la plus probable *a priori*. C'est précisément ce que nous observons dans ces conditions : les participants tendent à favoriser les intentions qui

⁸ Cela s'applique également à l'expérience de pensée de Jacob & Jeannerod (2005), que nous mentionnions en introduction : en l'absence d'attentes particulières sur la personnalité de celui qui tient le scalpel (Mr Hyde ? Dr Jeckyll ?), « manier un scalpel » peut indifféremment dénoter l'acte de torturer et celui de soigner.

satisfont leurs attentes préalables, et à se défier de l'information perceptive associée à la scène d'action.

Enfin, les conditions « sociales » se caractérisent par une contamination de ces attentes *a priori* généralisée à tous les niveaux d'information visuelle, même élevé. Nous expliquons ce recours massifs aux *a priori* par les propriétés de la situation présentée aux participants (Chambon et al., ci-dessus). Les contextes d'interaction entre agents sont en effet connues pour amorcer des attentes modulaires, social-spécifiques, qui contribuent – avant même que l'information visuelle ne soit traitée, ou en parallèle de ce traitement – à favoriser certaines causes intentionnelles (e.g. le second joueur coopère si le premier joueur a coopéré / il fait défaut si le second joueur a fait défaut) au détriment de leurs alternatives (coopère toujours, fait toujours défaut, sans stratégie particulière). Cette préférence précoce en direction d'un certain mode de réciprocité (ici, le mode 'coup-pour-coup'), et la persistance de cette préférence en condition d'information visuelle élevée, pourrait suggérer l'existence d'un système *pré-traitant* l'information de contexte perceptif. Ce système déterminerait la nature de la situation rencontrée en vertu de ses propriétés perçues : l'action observée est-elle intentionnelle ? la situation est-elle sociale ? L'existence d'un système de pré-traitement de ce type expliquerait dès lors pourquoi, lorsque l'action est identifiée comme se déroulant au sein d'un contexte d'interaction, les participants manifestent des attentes précoces sur la manière dont les agents sont, dans ce contexte, susceptibles d'interagir. La dépendance accrue des participants à leurs *a priori* semble ici traduire une difficulté à se désengager de ces intuitions domaine-spécifiques.

La mise en évidence d'un biais *a priori* en faveur du mode de réciprocité 'coup-pour-coup' invite à formuler des hypothèses sur les bases évolutives de cette préférence. Les situations d'interaction sont, par défaut (même si elles ne le sont pas toujours nécessairement), des situations où chacun des participants tend à maximiser les bénéfices de l'interaction, tout en veillant à protéger sa 'bonne' réputation. Il a été montré que ce genre de situation tendait naturellement à exclure les tricheurs (« always defect ») ou les individus trop altruistes (« always cooperate ») pour converger vers un équilibre caractérisé par l'adoption de la stratégie « coup-pour-coup » (Rapoport & Chammah, 1965; André & Day, 2007). Le 'coup-pour-coup' est en effet le modèle qui explique le mieux comment la coopération peut émerger entre des individus mus par des intérêts égoïstes, et comment les systèmes coopératifs résistent aux tricheurs en les excluant, de proche en proche, de la coopération (Axelrod, 1997). Puisque le mode du 'coup-pour-coup' est le mode de réciprocité qui garantit, au mieux, la maximisation des gains à long terme, ce mode a pu être sélectionné en raison de son avantage sélectif particulier. Il n'est donc pas à exclure que les situations d'interaction sociale, lorsqu'elles sont identifiées comme telle, amorcent *par défaut* ce mode de réciprocité optimale. Cette préférence précoce, et peut-être irrépressible, pour le régime du 'coup-pour-coup' expliquerait dès lors pourquoi les sujets de la présente étude manifestent une dépendance accrue à leurs *a priori* (*i.e.* à ce mode de réciprocité) lorsque l'action de l'agent est précisément identifiée comme s'accomplissant dans un contexte d'interaction supposé *par défaut* mutuellement avantageux (*i.e.* un contexte où chaque joueur essaie de maximiser ses gains tout en préservant sa réputation).

5.2 Mentaliser sous influence : le cas de la schizophrénie

En introduction, nous avons formulé l'hypothèse que les troubles de la mentalisation rapportés dans la schizophrénie devraient se traduire sous la forme d'une dépendance anormale *i)* aux informations perceptives véhiculées par la cinématique de l'action observée,

sinon *ii*) aux connaissances préalables que le patient entretient sur le type d'action présenté. Nous nous attendions également à ce que la forme de cette dépendance varie selon la dimension clinique (négative ou productive) prévalant chez le patient.

Les quatre groupes de patients que nous avons testés présentent des patterns de réponses relativement similaires à celles des participants contrôles. En revanche, des profils de performance assez nets émergent lorsque l'on corrèle le type de réponse produite avec le profil symptomatologique des patients. En particulier, nous avons pu mettre en évidence que la sévérité de la symptomatologie productive était étroitement associée à une dépendance accrue aux informations *a priori*. Les patients souffrant d'hallucinations, de délires ou de troubles de la pensée formelle présentent en effet une défiance significative à l'encontre des informations perceptives – exogènes –, et une confiance accrue à l'égard des leurs attentes préalables (*prior expectations*). A l'inverse, la sévérité de la symptomatologie négative se trouve associée à une sensibilité accrue aux informations perceptives et à une absence de référence aux informations endogènes – *a priori*.

Enfin, ces profils de performance sont également sensibles aux variations de complexité intentionnelle. Les patients « productifs » manifestent en effet une dépendance prononcée à l'égard des informations *a priori* dans les conditions « superordonnées », conditions dans lesquelles, nous l'avons vu, l'ambiguïté perceptive incite à répondre en direction des intentions dont la probabilité *a priori* est la plus élevée. Les patients déficitaires sont, quant à eux, plus attentifs aux informations perceptives dans les conditions « sociales » et tendent à se défier des informations *a priori* dont ils disposent quant au mode d'interaction privilégié par la séquence (*i.e.* le mode 'coup-pour-coup'). Dans la littérature, l'hétérogénéité des performances pourrait être redevable de ce pattern de dépendance spécifique, variable selon la dimension clinique du patient et accentué par les caractéristiques (sociale ou non-sociale, motrice ou superordonnée) de l'intention ciblée.

Au total, l'hypothèse selon laquelle un déficit en mentalisation doit être exploré au niveau de ses mécanismes les plus élémentaires, est séduisante à plusieurs titres. Elle est parcimonieuse, d'une part, puisque les processus qu'elle incrimine ne sont pas spécifiques des troubles de la mentalisation et débordent même le champ des cognitions sociales ; or, la schizophrénie est bien loin d'être une pathologie réductible à un dysfonctionnement des processus de Théorie de l'Esprit. Ce défaut d'accès aux états mentaux d'autrui, s'il est avéré, s'inscrit en effet sur un arrière-plan de perturbations diffuses, en partie aspécifiques (pour une revue des perturbations cognitives recensées à ce jour dans la schizophrénie, voir Barch, 2005 ; Mesholam-Gately et al., 2009). L'hypothèse d'une interaction inappropriée entre vraisemblance perceptive et informations *a priori* a donc le mérite d'être assez élémentaire pour embrasser le fonds de ce déficit généralisé.

Cette hypothèse, d'autre part, est particulièrement cohérente avec les données de la littérature. Il a été montré à maintes reprises que les patients déficitaires accordaient un privilège excessif aux informations sensorielles – exogènes, c'est-à-dire prenant source dans l'espace extra-corporel. Chez les patients alexithymiques, en particulier, on relève une tendance marquée à se concentrer sur les événements extérieurs plutôt que sur les expériences intérieures (Taylor, 1994). Cette sensibilité excessive aux informations exogènes pourrait témoigner plus généralement du caractère stimulo-induit des comportements déficitaires ; elle expliquerait également la tendance au repli des patients souffrant de symptômes négatifs ainsi que l'ensemble des troubles de l'initiation volontaire fréquemment rapportés dans cette dimension particulière de la symptomatologie (Frith, 1992). Le profil inverse, en revanche, pourrait prévaloir chez les patients dits « productifs ». Les patients souffrant de délires, en particulier, présentent une vulnérabilité

particulière aux inférences arbitraires. Dans certaines tâches de raisonnement probabiliste, ces patients produisent significativement plus de conclusions hâtives (*jump-to-conclusion*) que des sujets témoins appariés en âge et niveau d'éducation (Brankovic & Paunovic, 1999 ; Colbert et al., 2002). Il a été suggéré que cette tendance à conclure hâtivement, sur la foi de peu d'informations exogènes, traduisait une confiance excessive à l'égard des informations *a priori*, auto-générées, et sur la base desquelles le patient aurait tendance à faire porter tout le poids de sa décision (Jones, 1999). Le tableau clinique de la croyance délirante présente d'ailleurs le même type d'« asymétrie cartésienne » : les informations qui surviennent dans l'environnement local ne sont pas traitées avec le même degré de confiance que les informations endogènes, pour leur part soutenue avec un degré de conviction extrême (Huq et al., 1988 ; McCabe, 2004). Cette asymétrie expliquerait ainsi le caractère imperméable de la croyance délirante, dont le contenu n'est souvent pas ou peu révisé, en dépit des preuves empiriques qui le contredisent en permanence.

5.3 Un modèle hiérarchique des représentations d'actions intentionnelles : système-miroir et *predictive coding*

Comme suggéré plus haut, la mise en évidence d'une interaction entre informations visuelles et informations *a priori*, et d'une variation de cette interaction selon le niveau intentionnel ciblé, invite à privilégier un modèle *hybride* de la compréhension de l'action – mobilisant tantôt des mécanismes simulationnistes, tantôt des mécanismes inférentiels opérant sur des connaissances domaine-spécifiques de plus haut niveau. Un nombre croissant d'observations empiriques et de travaux théoriques suggèrent que ces deux classes de mécanismes pourraient jouer un rôle complémentaire dans les processus de mentalisation (Mitchell, 2005; Keyzers & Gazzola, 2007; Brass et al., 2007; de Lange et al., 2008). Les résultats de la présente étude plaident en faveur d'un modèle hybride de ce type : dans ce cadre, l'observateur mobiliserait l'une ou l'autre classe de mécanismes selon que l'information sensorielle disponible est, ou n'est pas assez fiable pour amorcer un processus de simulation.

Récemment, Kilner et collaborateurs (2007a, 2007b) ont proposé de rendre compte de l'interaction de ces deux sources d'informations – *a priori* et perceptive – au sein d'un modèle hiérarchique de la compréhension de l'action. Le formalisme employé, celui du codage prédictif (*predictive coding*), est ici particulièrement approprié pour rendre compte de la manière dont un observateur résout le *problème inverse* de la perception intentionnelle (*i.e.* lorsque plusieurs intentions sont possibles sur la foi de ce qui est observé). Parce qu'il est biologiquement plausible, ce modèle permet en outre de faire des prédictions relativement fines sur la nature des mécanismes cérébraux impliqués.

Dans le cadre de ce modèle, il est postulé que des régions cérébrales hiérarchiquement distinctes (et possédant, ou non, des propriétés « miroir ») prennent en charge les différents niveaux de compréhension de l'action: 1) le niveau, général est descriptif, de l'« intention », 2) celui du but pour réaliser cette intention, 3) le niveau subordonné de la commande motrice et enfin, au pied de la hiérarchie, 4) le niveau de la cinématique du mouvement en cours d'exécution (voir également Grafton & Hamilton, 2007). Le formalisme du codage prédictif postule que chaque niveau de la hiérarchie formule des prédictions sur ce que devraient être les représentations du niveau inférieur, puis les prédictions du niveau supérieur sont comparées aux prédictions actuelles du niveau inférieur. Ainsi, étant donné la nature du but qu'il anticipe, le sujet prédira une commande motrice congruente avec ce but, et, sur la base de son propre système moteur, attendra par conséquent une certaine cinématique,

Les résultats de la présente étude peuvent être interprétés de manière cohérente à la lumière d'un modèle hiérarchique de ce type. L'intention motrice, telle que nous l'avons opérationnalisée, peut être directement prédite depuis l'observation de l'acte moteur lui-même, lorsque l'information visuelle est suffisamment fiable pour amorcer une comparaison avec la cinématique du mouvement prédite aux niveaux superordonnés. Dans ce cas, les performances des participants devraient être largement dépendantes de la minimisation de l'erreur de prédiction qui émerge de cette comparaison. Naturellement, cette comparaison est elle-même dépendante de la fiabilité de cette cinématique : lorsque l'information visuelle est trop incomplète, aucune comparaison ne peut avoir lieu, et de fait, les niveaux inférieurs ne peuvent ajuster leur représentation aux estimations superordonnées de la hiérarchie. La comparaison, suggère le modèle de Kilner, est ici impossible car l'erreur de prédiction est trop large : la cinématique observée sous-détermine en effet largement l'ensemble de ses causes possibles au niveau superordonné.

Dans la présente étude, l'impossibilité de cette comparaison se traduit chez les participants par une tendance à répondre en direction de l'intention qu'ils estiment *a priori* (*i.e.* avant l'observation de la séquence elle-même) la plus probable. Cette sur-dépendance aux *a priori* pourrait témoigner de l'existence d'un circuit de boucles récursives entre les niveaux inférieurs et supérieurs de la hiérarchie corticale. Ces boucles récursives pourraient être mobilisées lorsque l'information perceptive est trop incomplète pour engager une comparaison entre cinématique observée et cinématique attendue, ou, pour le dire autrement, lorsque miser par défaut sur nos attentes préalables est une meilleure garantie pour le succès de la prédiction – même si ces attentes peuvent, à l'occasion, contredire l'évidence perceptive.

5.4 Du système-miroir au réseau de la Théorie de l'Esprit (ToM network) : les bases neurales du *mindreading*

Le modèle hiérarchique de Kilner est particulièrement intéressant pour notre propos car il intègre à la fois des régions connues pour avoir des propriétés miroir (*lobule pariétal inférieur* : Fogassi et al., 2005 ; *cortex prémoteur* : Umiltà et al., 2001 ; Kilner et al., 2004) et des régions situées hors du système miroir humain (*sulcus temporal supérieur postérieur* : German et al., 2004 ; *cortex préfrontal médian* : Gallagher & Frith, 2003 ; Amodio & Frith, 2006). De fait, ce modèle autorise des prédictions assez fines sur les régions qui pourraient être impliquées dans l'attribution d'intentions – motrices, superordonnées ou sociales. Nous avons vu en effet que connaissances *a priori* et informations perceptives interagissaient au cours du processus de reconnaissance intentionnelle, et que la forme de cette interaction dépendait du type d'intention testé. Cette différence – dans le privilège accordé tantôt aux informations *a priori*, tantôt aux informations perceptives – devrait se traduire par l'engagement de régions cérébrales spécifiques.

Certaines régions sont en effet connues pour traiter préférentiellement les informations perceptives associées au mouvement biologique : c'est le cas du (1) sulcus intra-pariétal antérieur (AIPs), (2) de la partie postérieure du sulcus temporal supérieur (pSTS) et (3) du cortex prémoteur latéral :

1. AIPs : il est depuis longtemps admis que AIPs (Anterior Intra-Parietal sulcus) est sensible aux mouvements de saisie (*grasping*) visuellement guidés (Faillenot et al., 1997). Cette région est probablement l'homologue chez le singe de l'aire intra-pariétale (AIP), dont les neurones sont connus pour répondre à la fois aux composantes motrice et visuelle de la saisie manuelle (Sakata et al, 1995). AIPs est

également recruté par l'observation des mouvements d'autrui (Bonda et al., 1996) et, plus récemment, il a été suggéré que cette région était sensible au but de l'action (Hamilton & Grafton, 2006).

2. pSTS : le Sulcus Temporal Supérieur, dans sa partie postérieure, est une zone de convergence multimodale (Barnes & Pandya, 1992) recrutée lorsque les sujets observent les mouvements d'agents biologiques (Decety & Grèzes, 1998 ; Puce & Perrett, 2003). Au sein du pSTS pourrait siéger une dissociation fonctionnelle entre une sous-région qui contribuerait à différencier les mouvements biologiques des mouvements non-biologiques, et une seconde sous-région plus postérieure et inférieure qui répondrait indifféremment aux mouvements biologiques et non-biologiques (Pelphrey et al., 2003). Cette région est également sensible à certains aspects du contexte dans lequel le mouvement biologique est perçu (Pelphrey et al., 2003). Un nombre croissant d'études suggère néanmoins que le rôle du pSTS dans le codage des mouvements biologiques est une hypothèse trop réductrice, qui ne rend pas pleinement compte de la contribution de cette structure aux processus de reconnaissance intentionnelle en général. En effet, le pSTS est également sensible à la relation entre le mouvement observé et la structure de l'environnement, et coderait pour les séquences de mouvements dirigés vers un but présent dans l'environnement immédiat du sujet (German et al., 2004 ; Saxe et al, 2004).
3. LPMc : il est aujourd'hui admis que le cortex pré-moteur latéral (Lateral Pre-Motor cortex) joue un rôle central dans la planification de l'action. Plusieurs études ont également mis en évidence l'implication de ses parties dorsal et ventral dans l'exécution et l'observation d'actions intentionnelles (Grèzes & Decety, 2001). Sur la foi de ces observations, il a été suggéré que le LPMc appartiendrait au système-miroir humain (Iacoboni et al, 1999 ; Iacoboni et al, 2005 ; Buccino et al, 2001). Comme l'aire intra-pariétale, le LPMc serait recruté lorsque le sujet exécute ou perçoit des mouvements articulés de segments corporels, indépendamment de leurs formats sensoriels, visuels ou auditifs (voir Overwalle & Baetens, 2009, pour revue).

Ces différentes régions (STS en particulier) pourraient naturellement être sensibles aux niveaux d'information visuelle que nous manipulons dans les 4 tâches de notre étude, et représenter les preuves sensorielles que l'observateur, durant la reconnaissance, accumule jusqu'à la décision finale. L'interaction que nous observons entre information visuelle et *a priori* suggère en outre que ces représentations visuomotrices de bas niveau pourraient subir l'influence modulatoire de régions superordonnées dans la hiérarchie corticale. Parmi ces régions, le cortex cingulaire antérieure dorsal (dACC) et le cortex préfrontal ventromédian (vmPFC) en particulier retiendront notre attention.

Mindreading : dACC, vmPFC et templates intentionnels

Si le cortex prémoteur latéral, l'aire intrapariétale et le sulcus temporal supérieur sont liées entre eux par des connexions réciproques (*cortex prémoteur/cortex préfrontal* : Luppino et al., 1999 ; *STS/cortex préfrontal* : Seltzer & Pandya, 1994 ; Rushworth et al., 2006), ils entretiennent également des relations fonctionnelles avec des régions plus antérieures, dont le cortex préfrontal médian (mPFC) (Morecraft et al., 2004). Le mPFC – dont la partie antérieure du gyrus paracingulaire –, nous intéresse particulièrement dans la mesure où de nombreuses études pointent son implication dans une gamme variée de tâches de mentalisation, incluant la reconnaissance d'intentions superordonnées (Walter et al., 2004) et d'intentions sociales, de nature coopérative ou compétitive (Ciaramidaro et al., 2007, Gallagher & Frith, 2003 ; Amodio & Frith, 2006, pour revue). En règle générale, le mPFC

est recruté lorsque les participants doivent considérer les états mentaux d'un tiers, seul ou en interaction, indépendamment du type de stimulus présenté (image, vidéo, histoire écrite ou verbale) ou des caractéristiques de l'instruction (Frith & Frith, 2003, 2006).

Que l'on retrouve de manière cohérente le mPFC dans les tâches requérant la lecture d'intentions superordonnées ou sociales (séquence articulée de sous-buts, réalisés seul ou en interaction avec un autre agent) suggère l'implication de cette région dans l'intégration de l'information perceptive courante avec 1) le but distant qui gouverne la séquence d'actes observée, et 2) les attentes préalables des participants sur le but visé par la séquence, formulées en vertu de leur expérience passée de ce type de séquence. Dans une étude récente, C. Summerfield et E. Koechlin observent en effet un recrutement du cortex préfrontal ventro-médian (vmPFC) lorsque les participants doivent appairer les informations perceptives d'une situation donnée avec le modèle *a priori* (*template*) qu'ils ont formé de cette situation (Summerfield & Koechlin, 2008). Le signal d'appariement, lorsqu'il survient, renforce les prédictions initiales du sujet et l'encourage à persévérer en direction du modèle attendu (voir également Koechlin et al., 2002). Cette observation est particulièrement importante, car, dans les tâches que nous proposons, elle s'ajuste parfaitement à la tendance qu'ont les participants à *persévérer* en direction du biais qui leur a été imprimé, c'est-à-dire à répondre préférentiellement en direction des intentions dont la probabilité *a priori* est la plus élevée.

Certaines régions du cortex préfrontal médian pourraient jouer donc un rôle central dans la représentation, la maintenance ou la production des attentes *a priori* que les participants de notre étude manifestent pour les intentions en direction desquelles ils ont été biaisés⁹. Parmi ces régions, la partie rostrale antérieure du cortex cingulaire (dACC) mérite une attention particulière : dans une tâche de contrôle cognitif sous-tendue par une composante motivationnelle, Kouneiher et collaborateurs ont en effet observé que le dACC avait la propriété de potentialiser (*energetize*) les structures préfrontales latérales en vertu des demandes de la tâche et des événements passés enregistrés par le sujet (Kouneiher et al., 2009). Dans la série de tâches que nous avons présentées ci-dessus, l'observation d'une interaction entre fiabilité de l'input visuel et attentes préalables des participants suggère un mécanisme similaire de « potentialisation » de l'information visuomotrice. Le cortex cingulaire antérieur pourrait jouer un rôle central dans ce processus modulateur *top-down* : en présence de données perceptives incomplètes, le dACC favoriserait la prédiction, en relevant, sur la foi des attentes préalables de l'observateur, le niveau d'activation *a priori* des populations neuronales dédiées (« bonus au démarrage » ; Summerfield & Koechlin, 2008 ; Domenech & Dreher, en révision).

5.5 Etude IRMf : problématique générale

⁹ Cette interprétation ne signifie pas nécessairement que le système miroir ne joue pas un rôle dans la représentation de ces attentes *a priori* (*priors*). Il a par exemple été suggéré que le système miroir était efficace (voire évolutionnairement adapté) pour la prédiction des états subséquents de séquences d'actions, une fois ces actions étant reconnues comme dirigées vers un but. Cette fonction *prédictive* des neurones-miroir pourrait participer à la compréhension, non seulement du « quoi » (*what*) de l'action (*quel mouvement* le sujet met-il en œuvre pour accomplir son intention), mais également du « pourquoi » (*why*) de l'action (pour la réalisation de *quel but* le sujet met-il en œuvre ce mouvement). Cette hypothèse présuppose néanmoins que la relation entre l'action et son effet (le but) soit connue à l'avance de l'observateur et qu'elle ne varie pas, ou peu (*familiar action-effect association*, Rizzolatti & Craighero, 2004 ; Csibra & Gergely, 2007). Or, l'utilisation de séquences commutatives (actions dont la position est interchangeable au sein de la séquence) devrait ici limiter le rôle potentiel du système-miroir dans la prédiction des intentions superordonnées que nous manipulons. Dans cette condition, la référence accrue des participants à leurs *priors* pourrait d'ailleurs indiquer une manière de suppléer l'absence de prédictabilité interne des séquences.

L'étude IRMf que nous avons conduit, en collaboration avec Guillaume Barbalat, Philippe Domenech et Etienne Koechlin, vise précisément à déterminer le rôle des régions mentionnées ci-dessus, dans 2 tâches adaptées de la série de protocoles que nous avons développés dans l'étude comportementale précédente. Trois questions nous intéressent principalement :

1. La première d'entre elles concerne l'identité des régions cérébrales dont l'activité pourrait être modulée par les attentes a priori des participants sur les intentions guidant les actions, ou séquences d'actions, qu'ils observent.
2. La seconde vise à déterminer la manière dont la « preuve » visuelle (basse, modérée, élevée ou très élevée) est, dans ces régions, accumulée jusqu'à la décision finale ; et, le cas échéant, comment le processus d'accumulation d'informations visuelles peut se trouver biaisé par les attentes préalables des participants (interaction prior*information-visuelle).
3. La troisième, enfin, a trait aux variations de cette influence selon le niveau intentionnel considéré – moteur, superordonné, social (interaction prior*information-visuelle*intention). Nous faisons l'hypothèse que la dépendance accrue aux priors observée dans les conditions superordonnée et sociale de notre étude comportementale devrait se traduire par un engagement distinct des régions qui représentent l'interaction priors*information-visuelle.

Dans les pages qui suivent, nous exposons la méthode générale de cette étude et les résultats préliminaires que nous avons obtenus sur un groupe de 18 sujets témoins non-psychiatriques.

Etude IRMf

Valérian Chambon, Guillaume Barbalat, Philippe Domenech, Etienne Koechlin, Elisabeth Pacherie & Chloé Farrer

(rapport préliminaire)

6 Comment le cerveau représente-t-il la variété des intentions ?

6.1 Participants

18 sujets ont accepté de participer à l'étude (9 hommes et 9 femmes, entre 22 et 59 ans, SD = 9.8. ; niveau moyen d'éducation = 12.1, SD = 1.4 ; latéralité manuelle droite = 0.89, SD=0.1), pour laquelle ils ont été rémunérés à hauteur de 60 euros par session. Aucun d'entre eux ne présentait d'antécédents psychiatriques connus. Tous les participants étaient droitiers. Les critères d'exclusion étaient les suivants : difficultés d'attention majeures, troubles de l'humeur, addictions, maladie ou traumatisme neurologique, claustrophobie, âge supérieur à 65 ans, traitement par thymorégulateur, par antidépresseur ou par benzodiazépine au long cours, présence de corps métalliques et/ou de prothèses. Après

133

description complète de l'étude, les sujets ont signé le consentement éclairé nécessaire à la passation de l'expérience (CCP, AFSSAPS n°B80631-60).

6.2 Description de la tâche (Figure 2)

Protocole comportemental

Pour des raisons pratiques (durée des sessions IRMf et fatigue des sujets), l'étude a été réalisée sur deux sessions distinctes. Chaque session comprenait 2 tâches :

1. La session NON-SOCIAL comportait une tâche MOTOR (intention motrice) et une tâche SUPRA (intention superordonnée)
2. La session SOCIAL comportait une tâche SINGLE (intention motrice) et une tâche DUAL (intention motrice sociale)

Dans les 4 tâches, des séquences d'actions filmées, de durées variables, étaient présentées aux participants. La tâche des sujets consistait à reconnaître l'action (tourner ou déplacer) réalisée par un acteur sur un ou plusieurs cubes rectangulaires, de taille égale, disposés devant lui. Les participants avaient pour instruction de répondre le plus rapidement et le plus précisément possible, en appuyant sur l'une des deux touches d'un boîtier-réponse. Les séquences vidéo présentées dans cette étude ont été adaptées de celles utilisées dans l'étude comportementale précédente (Chambon et *al.*, p.83) L'expérience a été conçue et présentée sous Presentation® 9.90 (<http://www.neurobs.com/>).

Stimuli et procédure

Chaque session débutait par un écran noir (12.5 s) pendant lequel il était demandé au sujet de se concentrer. Puis la première séquence débutait. Dans les 4 séquences (*runs*) de chaque session, les essais se déroulaient de la manière suivante :

1) séquences MOTOR, SUPRA et SINGLE :

Une croix de fixation apparaît au centre de l'écran pendant une durée qui varie entre 500 et 2500 ms. Puis la croix de fixation disparaît et la vidéo commence : la main d'un acteur apparaît sur une surface uniforme, disposée devant

- i) 3 cubes alignés en arc-de-cercle autour d'un cube central (MOTOR),
- ii) un agencement de trois cubes centraux, et un dernier cube disposé en bas à gauche de l'écran (SUPRA),
- iii) deux cubes disposés de part et d'autre d'une grille centrale (SINGLE).

La main reste immobile pendant une durée qui varie entre 700 et 1500 ms, puis l'acteur amorce un mouvement en direction du centre de la scène. La durée du mouvement varie elle aussi, sur quatre niveaux (calculés à partir du début du mouvement) : 1420 ms (EQ1), 1500 ms (EQ2), 1580 ms (EQ3), et 1780 ms (EQ4). Ces trois niveaux correspondent aux 4 niveaux d'information visuelle manipulés durant l'expérience et représentent 4 états d'achèvement variables de l'action en train d'être réalisée (tourner ou déplacer un cube). La vidéo interrompue est alors remplacée par un écran noir dont la durée varie entre 500 et 1000 ms (*jitter*), puis apparaît un écran-réponse avec deux lettres (T et D) correspondant aux réponses possibles que le sujet peut fournir (T pour tourner, D pour déplacer). L'ordre de présentation des lettres a été contrebalancé entre les participants. Une fois la réponse entrée, l'essai suivant débute selon les mêmes modalités (figure 10, p.218).

Les Temps de Réaction (TR) et les réponses des participants ont été récoltés pour chaque essai.

2) séquences DUAL :

Le décours de chaque essai est le même que pour les 3 trois tâches précédentes, à l'exception du début de la séquence vidéo : deux acteurs se tiennent face-à-face (en haut et en bas de l'écran), devant une grille centrale sur les bords de laquelle reposent deux cubes de taille identique. Les deux acteurs agissent tour-à-tour : le premier acteur (bas de l'écran) tourne ou déplace le cube situé à sa droite, puis le second acteur (haut de l'écran) tourne ou déplace le cube restant. La plage temporelle « main au repos » des 3 tâches précédentes est ici remplacée par l'action du premier acteur (bas de l'écran), dont la durée varie entre 1500 et 2000 ms. Puis le second acteur agit à son tour : comme précédemment, la vidéo est interrompue avant que l'action ne s'achève. La durée de l'action varie sur quatre niveaux qui correspondent aux 4 quantités d'informations visuelles manipulées durant la tâche (1420, 1500, 1580 et 1780 ms). L'essai suivant débute une fois que le sujet a donné sa réponse.

Les Temps de Réaction (TR) et les réponses des participants ont été récoltés pour chaque essai.

3) tâche contrôle :

Une tâche contrôle était introduite aléatoirement entre les essais des 4 études. La tâche consistait pour les participants à déterminer si deux cercles présentés avaient, ou non, la même couleur. Un premier cercle rouge apparaissait en haut et au centre de l'écran durant 500 ms, suivi par un second cercle, bleu ou rouge, situé en bas et au centre de l'écran. Le second cercle restait également 500 ms à l'écran, avant d'être remplacé par un écran noir (500-1000 ms), puis un écran-réponse sur lequel apparaissait deux lettres correspondant aux deux réponses possibles (O pour Oui, N pour Non). L'ordre de présentation des lettres était contrebalancé entre les participants, dont les réponses et les TRs ont été collectés pour chaque essai.

L'objectif de cette tâche était d'établir une *baseline* commune aux 4 études et de contrôler les variations en demandes attentionnelles spatiales liées aux mouvements de la main observée (du haut vers le bas de l'écran).

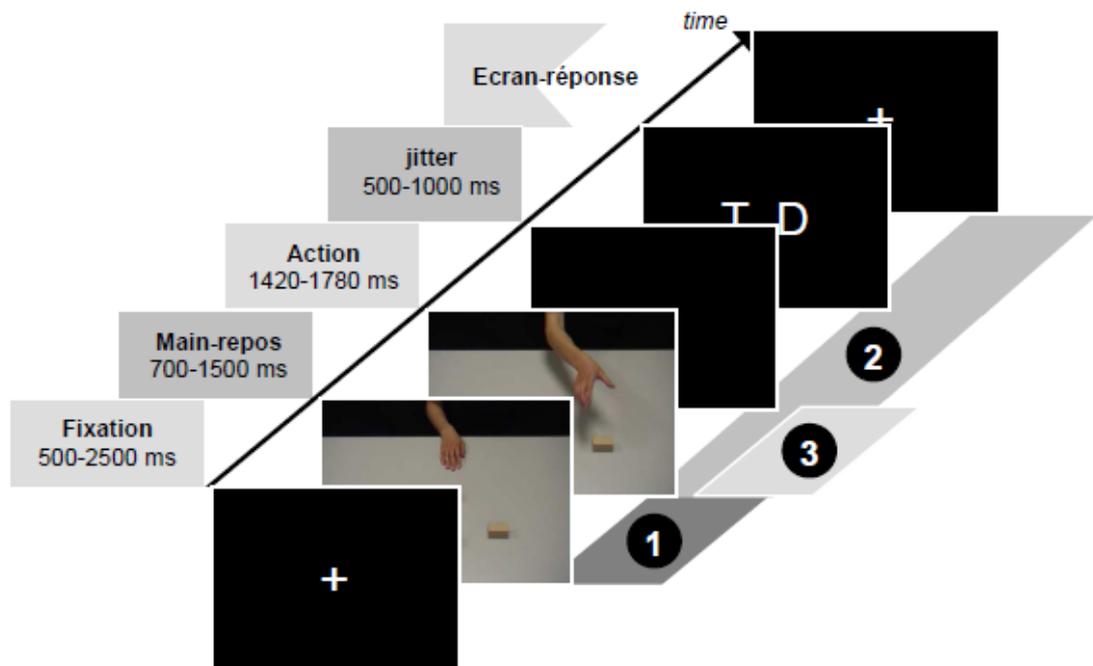


Figure 10. Déroulement temporel typique d'un essai (ici, pour la condition MOTOR). Chaque essai dure en moyenne 5s et s'interrompt une fois la réponse du sujet entrée. (1) Régresseur « preparation » ; (2) Régresseur « décision » ; (3) Régresseurs « evidence » {EQ1, EQ2, EQ3, EQ4}.

Instructions

Dans les séquences MOTOR et SINGLE, la tâche du sujet est de reconnaître l'action simple (tourner ou déplacer) finalement réalisée par l'acteur sur l'un des cubes disposés devant lui.

Dans les séquences SUPRA et DUAL, la tâche du sujet est de reconnaître cette action simple

1. sachant la forme centrale (Iso1, Iso2, Iso3) en passe d'être construite, i.e. l'intention superordonnée de la séquence de construction (tâche SUPRA)
2. sachant le coup (collaboratif ou compétitif) joué par le premier acteur au tour précédent, i.e. la stratégie sociale adoptée par le joueur précédent (tâche DUAL)

Entraînement

Avant les sessions IRMf, les sujets étaient entraînés à la tâche sur 2 séries (MOTOR / SUPRA ou SINGLE / DUAL) de 18 essais chacune. Cet entraînement était répété dans le scanner pendant l'acquisition structurelle (6 min).

Organisation des séquences

Afin d'amorcer chez les participants une préférence pour une action, forme ou stratégie particulière, 18 essais à très haut niveau d'information visuelle (1780 ms – EQ4) ouvraient chaque séquence expérimentale. Au cours de ces 18 essais, une action (MOTOR, SINGLE),

forme (SUPRA) ou stratégie sociale (DUAL) apparaissant plus souvent (2/3 – biais) que leurs alternatives (1/3 – non-biais). Cette procédure était ensuite répétée durant la séquence proprement dite.

Pour prévenir l'apparition de séries hautement prédictibles (successions prévisibles d'actions biaisées par exemple), chaque séquence était organisée en blocks de 9 essais chacun – comprenant 6 actions, formes, ou stratégies du même type (biais), et 3 autres d'un autre type (non-biais). L'ordre de présentation des essais était randomisé au sein de chaque block. Au total, chaque séquence comprenait 24 blocks de 9 essais chacun et chaque participant réalisait donc 432 essais par session (24*9*2). Ce total n'incluait pas les phases d'apprentissage et d'entraînement préalables à l'expérience proprement dite (hors scanner), ni les séquences de 18 essais à très haut niveau d'information ouvrant chaque séquence.

Acquisition et prétraitement des données IRMf

Les 18 sujets de l'étude ont été scannés dans un imageur 1.5 Tesla de marque Siemens (CERMEP, Bron). Pour chaque session, une acquisition structurale (image 3D pondérée en T1) était suivie de 4 séries d'acquisitions fonctionnelles (TR = 2.5 s, TE = 60 ms, matrice d'acquisition = 64x64 voxels{3*3*3 mm}, nombre de coupes par scan = 26, épaisseur de chaque coupe = 4 mm, distance entre chaque coupe = 0.4 mm). Les coupes ont été acquises dans un mode entrelacé de bas en haut.

Les stimuli étaient projetés sur un écran de sorte que les sujets pouvaient discerner parfaitement les séquences d'actions présentées. Les données ont été analysées à l'aide du logiciel SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>). Les cinq premiers volumes de chaque série ont été supprimés car le signal hémodynamique n'y était pas encore stabilisé. L'origine des images anatomiques et fonctionnelles a été réalignée sur la commissure antérieure.

Les prétraitements comprenaient des phases de réalignement (correction des artefacts dus aux mouvements de la tête), de recalage de l'image anatomique sur l'image fonctionnelle moyenne, de normalisation des images fonctionnelles dans l'espace du « template » du Montreal Neurological Institute (MNI), et de lissage spatial des données avec un filtre passe-bas gaussien, d'une largeur à mi-hauteur de 8 mm.

6.3 Analyses

La session DUAL (condition sociale) étant en cours de passation, nous ne présenterons ici que les analyses et les résultats de la session SINGLE (condition non-sociale).

6.3.1 Analyses des performances comportementales

Taux de réponses correctes. Nous avons réalisé des t-tests appariés afin de comparer entre eux les taux de réponses correctes pour les intentions 'biaisées' et 'non-biaisées' dans les conditions MOTOR et SUPRA.

Temps de réaction. Deux ANOVAs 2x3 à mesures répétées ont été réalisées sur les temps de réaction (TR) des participants, avec les facteurs Priors(biais vs. non-biais) et la quantité d'information visuelle en variables intra-sujets. Des tests post-hoc de Fisher ont été appliqués pour identifier les différences locales.

6.3.2 Données IRMf

Pour chaque condition (MOTOR et SUPRA), nous avons défini les régresseurs d'intérêt suivants (v. figure 10) :

1) Un régresseur '**Contrôle**' (boxcar modélisant les événements « Cercle haut » / « Cercle bas » durant la tâche contrôle)

2) Un régresseur « **preparation** » (boxcar modélisant l'événement « Main au repos » convolué avec la fonction de base hémodynamique).

3) Quatre régresseurs « **evidence** » {EQ1, EQ2, EQ3, EQ4} pour les 4 quantités d'informations manipulées (boxcar modélisant l'événement « Action »).

Pour les 4 quantités d'information visuelle {basse (Q1), modérée (Q2), élevée (Q3) et très élevée (Q4)}, la durée des régresseurs a été égalisée (1580 ms) de manière à obtenir une durée de stimulation visuelle identique pour des quantités d'informations différentes.

4) Un régresseur « **décision** » (du départ de l'action à la réponse du sujet).

5) Régresseur paramétrique '**priors**' : les régresseurs preparation et décision étaient modulés paramétriquement par le logarithme de la probabilité marginale de chaque intention, estimée sur la base des essais précédents. Cette valeur reflète la croyance actuelle du sujet sur la probabilité de survenue de chaque intention au début de l'essai en cours, et est mise à jour à l'issue de chaque nouvelle réponse.

L'hypothèse que les sujets utilisent de manière sans distinction *tout ce qu'ils ont vu* depuis le début de l'expérience jusqu'à l'essai en cours (*i.e.* qu'ils aient une mémoire infinie des événements passés – hypothèse de l'observateur Bayésien idéal) n'est cependant qu'une approximation qui, dans notre cas, ne décrit pas le comportement des participants de manière satisfaisante. C'est pourquoi nous avons fait l'hypothèse que la contribution de chaque événement passé aux croyances des participants décroissait au cours du temps, selon une fonction puissance. Nous avons donc estimé, pour chaque individu, le paramètre de cette perte d'information, et calculé le nombre d'essai nécessaire pour qu'un événement ne contribue plus qu'à hauteur de 5% de sa valeur informationnelle à l'établissement des croyances (horizon mnésique du participant) :

$$\log \left(\text{prob} (e_t = i) = \frac{\sum_{p=1}^t (\alpha_i^{-p} \times u(t)) + 1}{\sum_i \sum_{p=1}^t (\alpha_i^{-p} \times u(t)) + 1} \right)$$

où $u(t)$ est une fonction binaire qui prend 1 quand i se réalise à l'essai p , et 0 s'il ne se réalise pas (α est le paramètre de la fonction puissance et varie de 1 à 2 par un incrément dont la valeur est fixée par l'expérimentateur).

En condition MOTOR, nous obtenons un horizon mnésique moyen de 62 éléments ; en condition SUPRA, cet horizon est de 111 éléments (figure 11, ci-dessous). Les *priors* de

chaque participant ont été recalculés en vertu de cet horizon mnésique, puis entrés dans le régresseur paramétrique 'priors'.

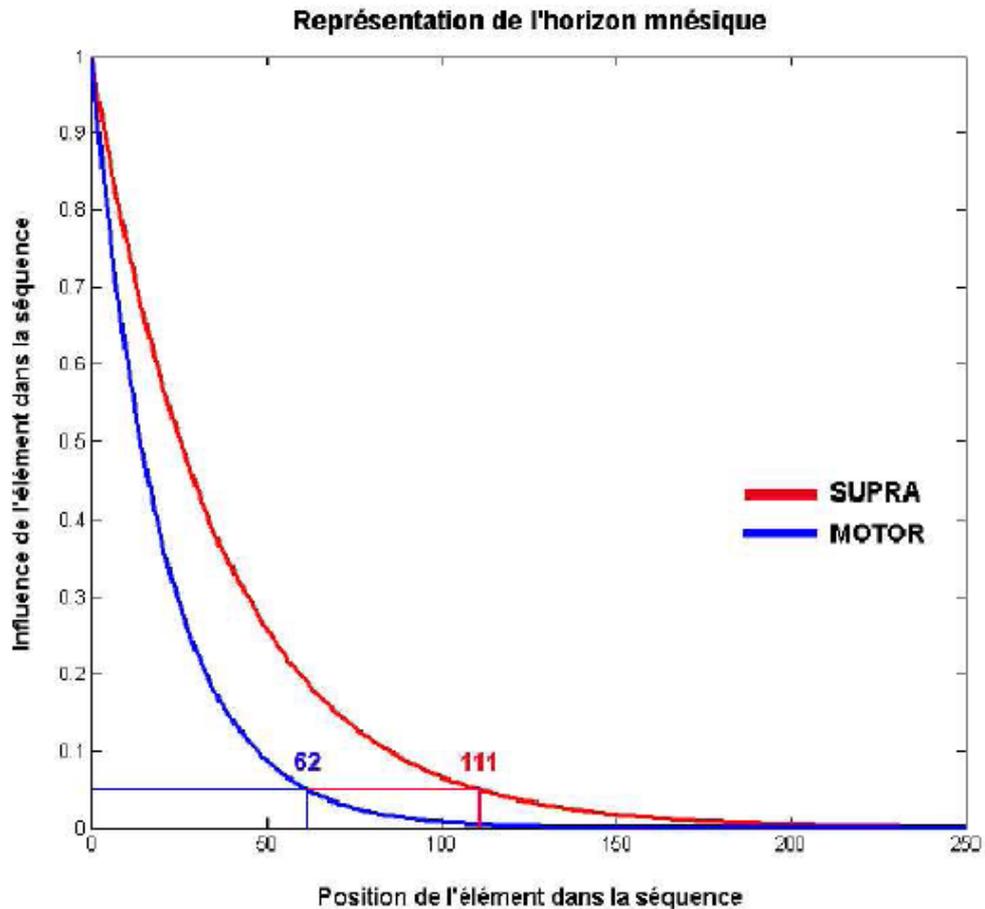


Figure 11. Représentation de l'horizon mnésique des participants dans les conditions MOTOR et SUPRA. En abscisse, la position de chaque élément relativement à l'élément actuel (origine). En ordonnée, l'influence de chaque élément sur la représentation des probabilités de la séquence : plus cet élément a d'influence, plus sa valeur est élevée. L'élément qui précède l'élément actuel est inclus dans le calcul de l'empan mnésique optimal tant que l'influence de cet élément est supérieure à 5% de l'influence maximale que cet élément peut revêtir.

6.3.3 Analyses des données IRMf

Analyse whole-brain . Nous avons dans un premier temps conduit des analyses exploratoires (« voxel-wise exploratory analyses ») au sein du cerveau entier. Un modèle linéaire général a été spécifié pour chaque sujet (analyses dites de « premier niveau »). Chaque modèle comprenait 4 sessions (2 x MOTOR, 2 x SUPRA), de 12 minutes chacune environ. Au sein de chaque session, les conditions « Main au repos », « Evidence »,

« Décision » et les régresseurs paramétriques correspondants ont été modélisées comme des variables indépendantes et le signal BOLD au sein de chaque voxel, comme une variable dépendante. La fonction de base canonique de la réponse hémodynamique classique (« hrf ») a été choisie pour modéliser le signal IRMf mesuré dans chaque voxel.

Nous avons, dans un second temps, conduit des analyses à effets aléatoires (« second niveau ») pour identifier, au niveau de groupe, les régions cérébrales dont l'activité co-variait avec les conditions décrites ci-dessus. Les activations sont reportées au seuil $p < 0.001$ non corrigé.

Pour identifier les régions dont l'activité variait linéairement avec la quantité d'information visuelle, nous avons défini une carte SPM{T}, produit d'une combinaison linéaire des 4 régresseurs « **evidence** », telle que :

$$\text{contraste}\{\mathbf{EVIDENCE}\} = (\text{EQ1} \times -6) + (\text{EQ2} \times -2) + (\text{EQ3} \times 2) + (\text{EQ4} \times 6)$$

Nous avons ensuite réalisés une analyse de conjonction entre les contrastes EVIDENCE{MOTOR} et {SUPRA} afin d'identifier les régions qui, dans ces deux conditions, s'activaient lorsque l'information visuelle augmentait linéairement. Nous avons également réalisé une analyse de conjonction entre les contrastes DECISION{MOTOR} et {SUPRA} afin d'identifier les régions qui, dans ces deux conditions, s'activaient lors du processus de décision intentionnelle.

Enfin, nous avons masqué (masquage exclusif, $p = .05$ *uncorrected*) les activations du régresseur « DECISION » en condition SUPRA par les activations de ce même régresseur en condition MOTOR (*SUPRA but not MOTOR*), et inversement (*MOTOR but not SUPRA*).

Analyses par régions d'intérêt. Enfin, en accord avec les hypothèses *a priori* formulées en introduction, nous avons conduit des analyses par régions d'intérêt dans les régions dont l'activité était modulée paramétriquement par le facteur 'priors', dans les conditions MOTOR et SUPRA. Ces régions incluent le gyrus préfrontal ventro-médian (BA 10), la partie médiane du cortex cingulaire (BA 24), et l'aire motrice pré-supplémentaire (pre-SMA).

Dans chacune de ces régions, nous avons cherché à déterminer si le signal IRMf (*percentage of signal change*) était influencé par le facteur 'information visuelle'. Les analyses ont été circonscrites à des sphères de 8 mm de rayon, centrées sur le *voxel pic*.

6.4 Résultats

6.4.1 Résultats comportementaux

Taux de réponses correctes (figure 12). En moyenne, les participants reconnaissent bien les intentions présentées, avec des performances supérieures à 87 % de réponses correctes sur l'ensemble des sessions. Comme attendu, les participants reconnaissent mieux les intentions à probabilités élevées ('Biais') que les intentions à probabilités diminuées ('Non-biais') (MOTOR : $t = 2.12$, $p < .05$; SUPRA : $t = 3.29$, $p < .005$). Les participants répondent également plus souvent en direction des intentions biaisées dans la condition SUPRA que dans la condition MOTOR ($t = 4.71$, $p < .001$).

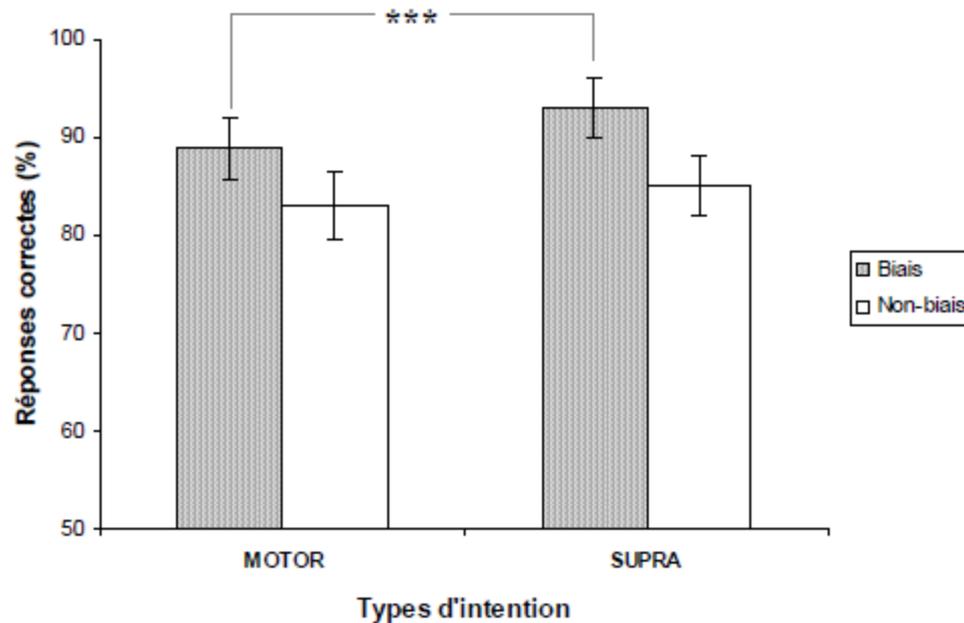


Figure 12. Taux de réponse correctes \pm SD pour les deux conditions MOTOR et SUPRA. En grisé, les réponses en direction des intentions avec la probabilité la plus élevée (« Bias »).

Temps de réaction (table 1). Comme attendu, les participants répondent plus rapidement en direction des intentions biaisées que non-biaisées, en MOTOR (effet principal 'priors' : $F(1,17)=15.42$, $p=.001$) comme en SUPRA (effet 'priors' : $F(1,17)=16.97$, $p<.001$). Les latences diminuent également avec l'augmentation de la quantité d'information visuelle dans les deux conditions intentionnelles (MOTOR : $F(2,34)=48.4$, $p<.001$; SUPRA : $F(2,34)=32.85$, $p<.001$). Enfin, le facteur 'priors' interagit significativement avec le facteur 'information visuelle' dans les deux conditions (MOTOR : $F(2,34)=3.42$, $p=.04$; SUPRA : $F(2,34)=4.1$, $p=.02$). Les participants répondent en effet plus rapidement en direction des intentions biaisées que non-biaisées lorsque l'information visuelle diminue. C'est vrai en particulier pour les quantités d'information basse (EQ1) et modérée (EQ2) (LSD Fisher : EQ1-b vs. EQ1-nb, all $p<.001$; EQ2-b vs. EQ2-nb, all $p<.05$; EQ3-b vs. EQ3-nb, all $p>.44$)

		Reaction Times (ms)		
Amounts of Visual Inf.		LOW	MODERATE	HIGH
EXPERIMENT				
Motor	Non-bais	649 \pm 164	554 \pm 144	421 \pm 121
	Bias	533 \pm 115	485 \pm 115	401 \pm 95
Supra	Non-bias	785 \pm 183	674 \pm 130	447 \pm 116
	Bias	623 \pm 137	548 \pm 133	432 \pm 109

Table 1. Temps de réaction \pm SD pour les trois niveaux d'information visuelle EQ1, EQ2 et EQ3. 'Bias' : intention dont la probabilité est la plus élevée.

6.4.2 Résultats IRMf (préliminaires)

6.4.2.1 Analyses exploratoires au niveau du voxel (« Exploratory voxelwise contrasts »)

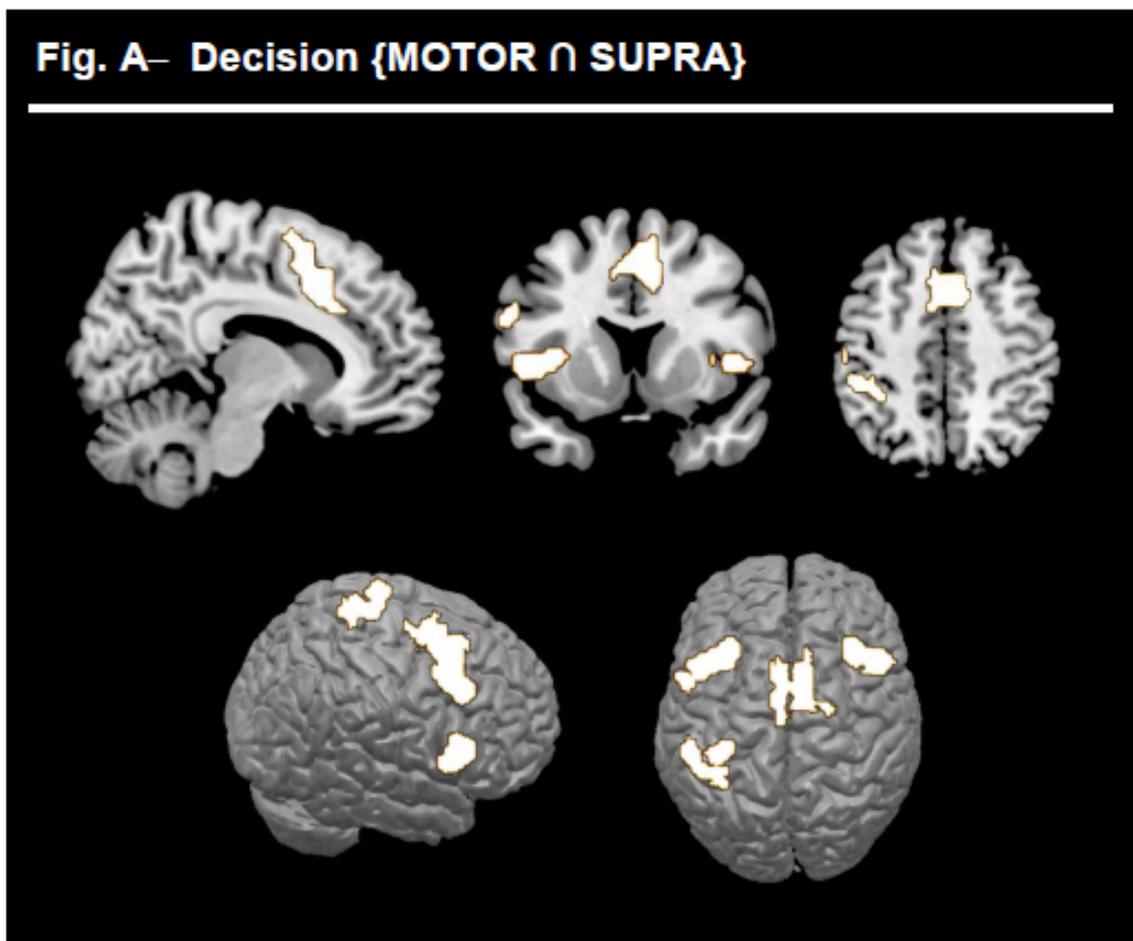
Preparation* priors (table 2)

En condition SUPRA, les régions dont l'activité varie linéairement avec les *priors* des participants ont été retrouvées dans la partie rostrale du cortex cingulaire (BA 32), le gyrus frontal supérieur médian droit (BA 9) et le gyrus cingulaire postérieur (BA 26/BA 23)

En condition MOTOR, les seules régions que nous retrouvons sont le cervelet et le thalamus gauche, mais l'activité dans ces régions n'atteint pas la significativité ($p=.06$).

Decision (table 2)

Dans les conditions MOTOR et SUPRA, un effet principal de la décision a été retrouvé dans l'insula bilatéralement, le gyrus frontal inférieur (BA 47/48) – bilatéralement également –, le sulcus intra-pariétal gauche (BA 2/3), le gyrus pariétal inférieur (BA 7), la partie rostrale du cortex cingulaire (BA 32), et le gyrus prémoteur latéral (BA 6/44) (**figure A, page suivante**).



[Fig. A]

Masquage exclusif {DECISION}

SUPRA but not MOTOR . Le masquage exclusif des activations SUPRA par les activations MOTOR révèle un engagement plus marqué du gyrus pariétal inférieur (BA 7) dans les conditions SUPRA, ainsi que le recrutement du gyrus frontal inférieur gauche (BA 45/48) et médian droit (BA 46/45).

MOTOR but not SUPRA . Un masquage exclusif des activations MOTOR par les activations SUPRA révèle, bilatéralement, le sulcus temporal supérieur (STS).

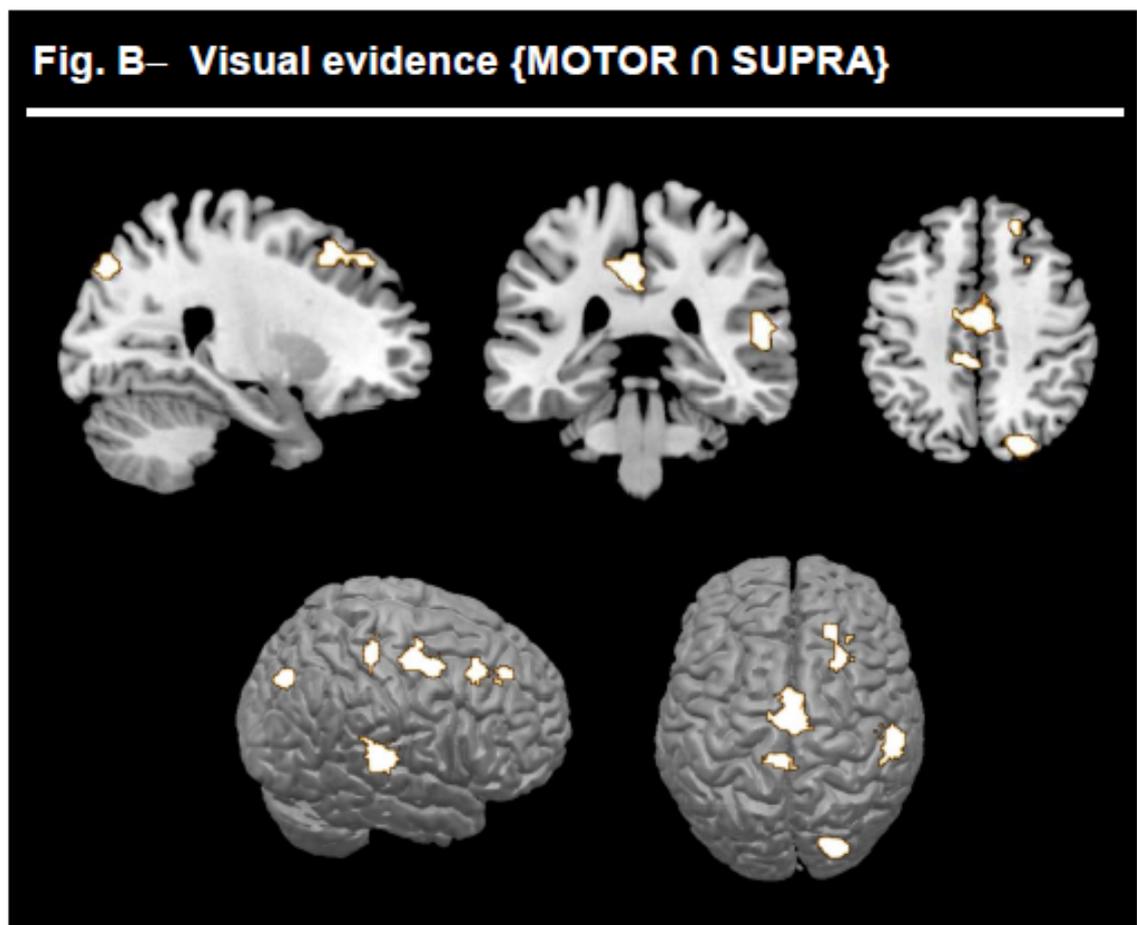
Decision*priors (table 2)

En condition MOTOR, les régions décisionnelles dont l'activité varie linéairement avec les *priors* des participants sont le gyrus occipital médian gauche (BA 18/19), le gyrus prémoteur (BA 6), le gyrus frontal médian supérieur (BA 10), la partie ventro-médiane du gyrus préfrontal (BA 10) et le gyrus cingulaire antérieur (BA 32).

En condition SUPRA, les régions décisionnelles dont l'activité varie linéairement avec les *priors* des participants sont le gyrus cingulaire médian dorsal (BA 24/6), le gyrus préfrontal ventro-médian droit (BA 10) et le gyrus cingulaire antérieur (BA 32).

Informations visuelles {EVIDENCE}

Les régions dont l'activité, dans les conditions MOTOR et SUPRA, varie linéairement avec l'information visuelle de cinématique ont été retrouvées dans le sulcus temporal supérieur droit (BA 22/42), la partie médiane du cortex cingulaire (BA 24), le gyrus cingulaire postérieur (BA 23), le gyrus frontal supérieur droit (BA 8/9), et le gyrus occipital supérieur droit (BA 19/BA 7) (**figure B, ci-dessous ; table 2**).



[Fig. B]

6.4.2.1 Analyses par régions d'intérêt fonctionnelles

L'activité dans le gyrus préfrontal ventro-médian droit démontre un effet principal de la condition (MOTOR > SUPRA, $F(1,17)=8.3$, $p=.01$), ainsi qu'un effet principal du facteur 'information visuelle' ($EQ1 < EQ2 < EQ3 < EQ4$, $F(3,51)=4.81$, $p=.005$). L'effet d'interaction entre la condition et le facteur 'information visuelle' est également significatif ($F(3,51)=2.72$, $p<.05$) : l'activité dans cette région est significativement modulée par les différentes quantités d'information visuelle dans la condition MOTOR seulement (**figure 13b, p.232**).

L'activité dans le gyrus préfrontal ventro-médian gauche démontre également un effet principal de la condition (MOTOR > SUPRA, $F(1,17)=12$, $p=.003$), ainsi qu'un effet principal du facteur 'information visuelle' (EQ1<EQ2<EQ3<EQ4, $F(3,51)=5.35$, $p=.002$). L'effet d'interaction condition*information visuelle' est seulement tendanciel ($F(3,51)=1.9$, $p=0.12$) : l'activité dans cette région tend à être modulée par les différentes quantités d'information visuelle dans la condition MOTOR seulement (**figure 14, p.233**).

L'activité dans l'aire motrice pré-supplémentaire démontre un effet principal de la condition (SUPRA > MOTOR, $F(1,17)>4.61$, $p>0.05$), indépendamment de la quantité d'information visuelle présentée ($F(3,51)<0.62$, $p>0.6$) (**figure 14, p.233**). L'activité dans la partie médiane du gyrus cingulaire démontre également un effet principal de la condition (SUPRA > MOTOR, $F(1,17)=4.61$, $p<.05$), et un effet principal du facteur 'information visuelle' (EQ1<EQ2<EQ3<EQ4, $F(3,51)=6.26$, $p=.001$). L'effet d'interaction condition*information visuelle' n'est pas significatif ($F(3,51)=.01$, $p=.89$) (**figure 13b, p.232**).

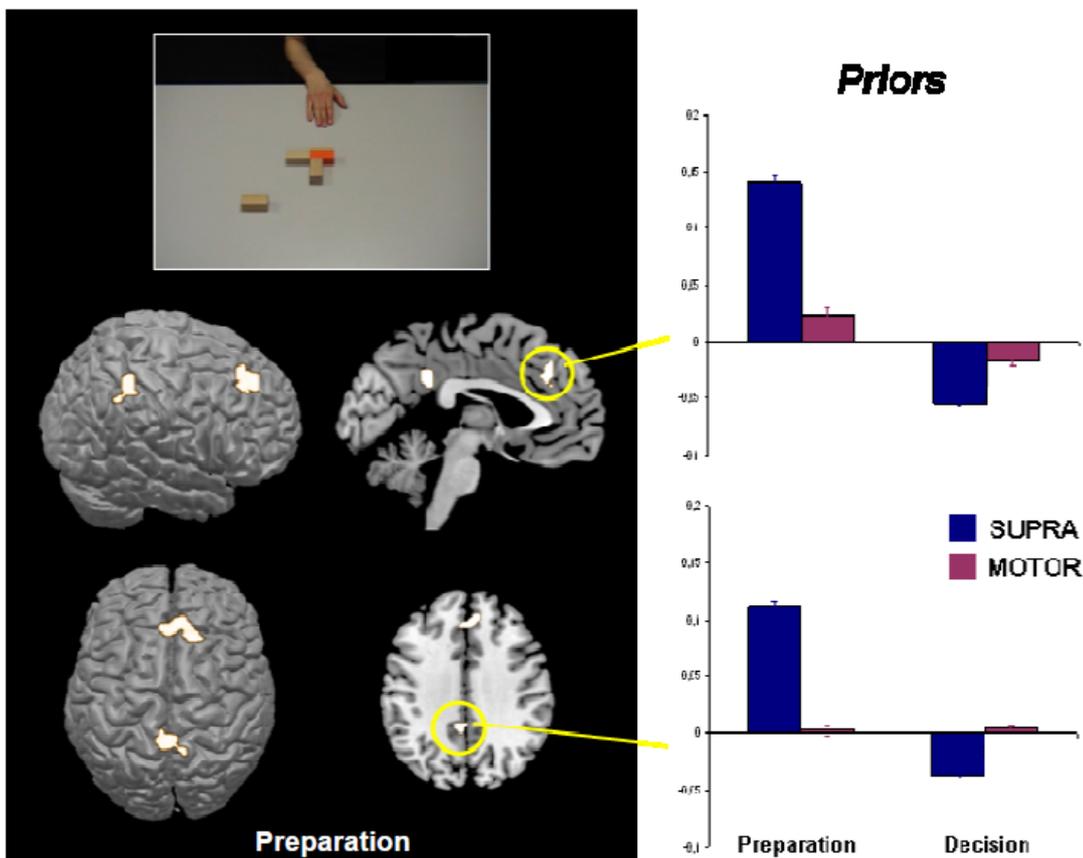


Figure 13a. Condition SUPRA : régions dont l'activité est modulée par les priors des participants durant la phase de préparation (« Main au repos ») (coefficients de régression moyennés \pm SEM).

A titre indicatif, nous avons extrait le pourcentage de changement de signal dans ces régions (gyrus cingulaire postérieur et gyrus cingulaire rostral) en réponse aux *priors* durant les phases de Préparation et de Décision, en conditions MOTOR (rouge) et SUPRA (bleu). En SUPRA, les régions dont l'activité covarie *positivement* avec les *priors* durant la Préparation tendent à covarier *négativement* avec les *priors* durant la phase de Décision.

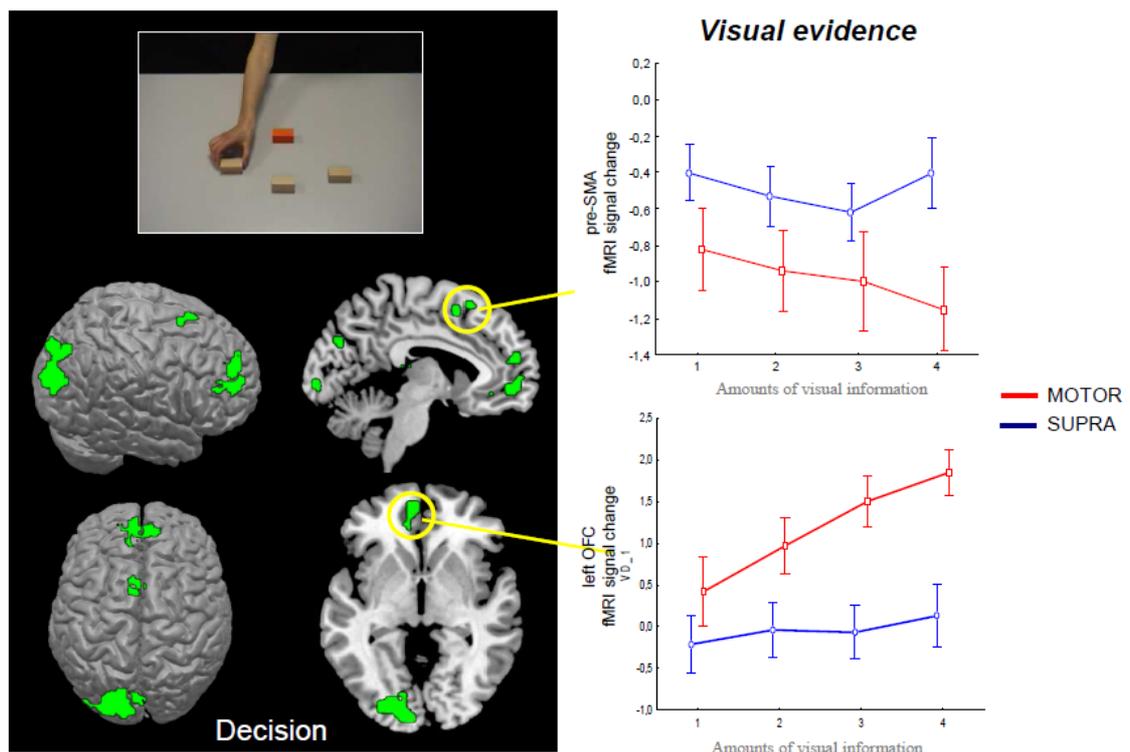


Figure 13b. Condition **SUPRA** : régions dont l'activité est modulée par les *priors* des participants durant la phase de **DECISION** (« Action »). Nous avons extrait le pourcentage de changement de signal dans ces régions (gyrus préfrontal ventro-médian (vmPFC) et gyrus cingulaire médian) pour les 4 quantités d'information visuelle (**Visual evidence**) manipulées dans les deux conditions **SUPRA** (bleu) et **MOTOR** (rouge). Le vmPFC est significativement modulé par l'information visuelle en condition **MOTOR** seulement.

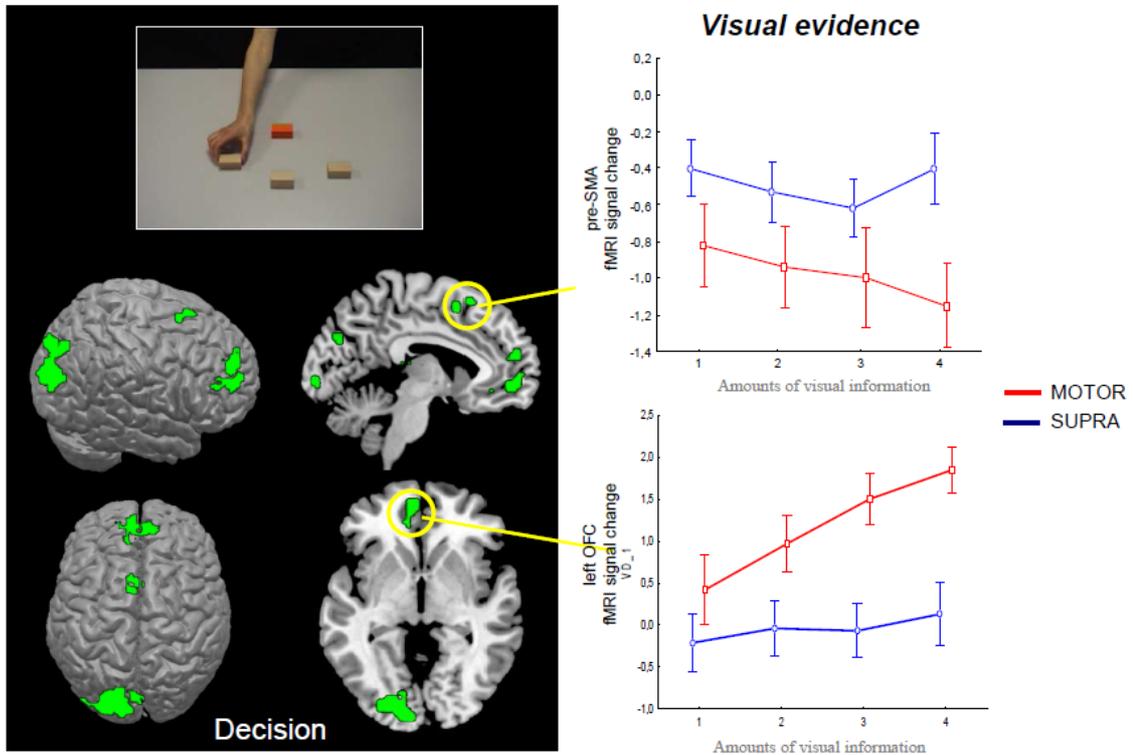


Figure 14. Condition **MOTOR** : régions dont l'activité est modulée par les priors des participants durant la phase de décision (« Action »). Nous avons extrait le pourcentage de changement de signal dans les régions du gyrus préfrontal ventro-médian (vmPFC) et dans la pré-SMA pour les 4 quantités d'information visuelle (**Visual evidence**) manipulées dans les deux conditions SUPRA (bleu) et MOTOR (rouge). Le vmPFC tend à être modulé par l'information visuelle en condition MOTOR seulement.

Table 2. Activations dans les régions cérébrales montrant un effet des facteurs 'priors' et 'visual evidence' durant les phases de PREPARATION et DECISION

Régions anatomiques	BA estimée	Coordonnées ^a			t Score	Volume (k _v)	p value*
		x	y	z			
PREPARATION+PRIORS							
MOTOR							
cervelet gauche	30	-15	-45	-21	4.98	18	.064
thalamus gauche	-	-24	-15	12	3.90	19	.038
SUPRA							
gyrus cingulaire postérieur	26	-6	-42	24	4.08	40	<.01
gyrus cingulaire rostral	32	6	39	39	4.44	75 ^b	<.001
gyrus frontal supérieur médian	9	-6	33	30	4.05	75 ^b	<.01
DECISION (MOTOR ∩ SUPRA)**							
insula gauche	-	-30	18	3	8.67	164 ^c	<.001
gyrus frontal inférieur gauche	48	-51	12	3	7.74	164 ^c	<.001
insula droit	-	33	24	3	7.01	143 ^d	<.001
gyrus frontal inférieur droit	45	48	15	3	6.98	143 ^d	<.001
gyrus cingulaire rostral	32	6	12	48	7.97	298 ^e	<.001
cortex prémoteur dorsal	6	6	6	60	7.42	298 ^e	<.001
gyrus pariétal inférieur gauche	40	-45	-39	48	7.11	109 ^f	<.001
gyrus postcentral	2	-39	-42	54	6.41	109 ^f	<.001
cortex prémoteur latéral	6/44	-57	9	24	7.80	35	<.001
gyrus pariétal supérieur gauche	7	-21	-69	42	4.91	77	<.001
DECISION+PRIORS							
MOTOR							
gyrus occipital médian gauche	18	-27	-90	21	5.65	246	<.001
cuneus	19	0	-81	30	4.09	32	.026
pré-SMA	6/8	-6	21	57	4.75	34 ^g	.022
SMA	6	-6	12	54	4.23	34 ^g	.022
gyrus préfrontal ventro-médian	10	-6	57	0	4.17	49	.008
SUPRA							
gyrus cingulaire médian	24	9	0	42	4.96	32	.019
gyrus préfrontal ventro-médian	10	6	48	-6	4.97	38	.011
EVIDENCE (MOTOR ∩ SUPRA)							
gyrus cingulaire médian	24	3	-9	42	4.64	118	<.001
gyrus cingulaire postérieur	23	-6	-33	42	3.74	39	.028
gyrus frontal supérieur droit	8/9	27	24	51	3.56	58	.01
sulcus temporal supérieur droit	22	54	-33	9	4.41	75	.004
cuneus	19	21	-78	42	4.08	48	.017

Abréviations : BA, Brodmann area. SMA : aire motrice supplémentaire

* cluster-level uncorrected

** Analyse de conjonction : pour des raisons de clarté, nous présentons ici les activations au seuil corrigé FWE

^aCoordonnées MNI

^bCes pics appartiennent au même cluster d'activation

^cCes pics appartiennent au même cluster d'activation

^dCes pics appartiennent au même cluster d'activation

^eCes pics appartiennent au même cluster d'activation

^fCes pics appartiennent au même cluster d'activation

^gCes pics appartiennent au même cluster d'activation

[Table 2.]

6.5 Discussion préliminaire – tâche SINGLE

Les deux tâches proposées ici invitaient les participants à identifier les intentions d'un agent interagissant avec des objets sans signification. D'une tâche à l'autre, nous faisons varier le rapport de l'action observée au but ou à l'intention que cette action réalisait. Dans la condition MOTOR, l'action dénotait directement le but sous-jacent (*one-to-one relation*) tandis que dans la condition SUPRA, l'action entretenait une relation équivoque avec l'intention qu'elle contribuait à réaliser (*one-to-many relation* : l'acte moteur « tourner », par exemple, pouvait aussi bien conduire à l'élaboration de la forme S1 que S2). Dans la première condition, les attentes des participants ('priors') portaient sur l'intention motrice de l'agent – c'est-à-dire un niveau de représentation sensori-moteur directement accessible à l'observation –, tandis que dans la seconde condition, les attentes portaient sur la forme construite par l'agent – c'est-à-dire un niveau de représentation partiellement abstrait des afférences visuo-motrices.

Les résultats que nous rapportons ici sont encore préliminaires et nécessitent, pour être discutés en toute rigueur, la conduite d'analyses supplémentaires. Plusieurs tendances se dégagent néanmoins, durant la phase de préparation comme durant la phase de décision proprement dite.

6.5.1 Une architecture cérébrale des représentations d'action

L'événement que nous avons modélisé comme phase de « décision » commence à l'initiation du mouvement et se termine lorsque le sujet donne sa réponse (on fait ici l'hypothèse que la phase durant laquelle le sujet accumule des « preuves » visuelles sur l'action en cours fait partie intégrante de sa décision finale). Dans les conditions MOTOR et SUPRA, le processus décisionnel recrute – indépendamment de la modulation des facteurs 'priors' et 'visual evidence' –, un réseau de régions pariétales et frontales similaire à celui que l'on retrouve habituellement dans les tâches de reconnaissance d'intentions. Ce réseau est particulièrement cohérent avec l'architecture cérébrale des représentations d'action décrite par Grafton et Hamilton (2007). Dans cette architecture, la compréhension du but est la conséquence d'un traitement hiérarchisé de l'action qui suppose, à chaque niveau de la hiérarchie, l'engagement des régions dédiées

- i) au traitement de la commande motrice (tendre le bras, ajuster la pince de saisie, etc.)
- ii) au traitement de l'action que cette commande réalise (atteindre ou saisir un objet),
- iii) à la reconnaissance du but proximal de l'action (tourner ou soulever un cube) et, le cas échéant,
- iv) à la compréhension du but *général* de cette action (tourner ou soulever un cube *pour* construire une forme particulière)

Ces différents niveaux de représentation de l'action intentionnelle recrutent plusieurs régions cérébrales distinctes que nous retrouvons dans la présente étude, et qui incluent :

1. le cortex pariétal supérieur postérieur, connu pour traiter les informations de cinématique bas-niveau (comme la trajectoire de la main ou la taille du grip),
2. le sulcus intra-pariétal antérieur gauche (AIPs), classiquement associé à l'observation des mouvements de saisie manuelle (Faillenot et al., 1997) et à la détection des interactions but/objet (« what », Hamilton & Grafton, 2006),
3. Le lobule pariétal inférieur (IPL), et
4. le gyrus frontal inférieur (IFG), tous deux sensibles au but général (« why ») de l'action, comme cela a pu être mis en évidence dans un paradigme récent de répétition-suppression (Hamilton & Grafton, 2008 ; voir également Iacoboni et al., 2005). Enfin,
5. le cortex prémoteur (PMC), dans ses parties dorsale et ventro-médiane, est souvent décrit comme un répertoire de représentations prémotrices recrutées à la fois par l'exécution et l'observation d'actions dirigées vers un but (Gallese et al., 2004). Le PMC jouerait à ce titre un rôle crucial dans l'accès aux intentions d'autrui, du moins lorsque ces intentions entretiennent des relations familières, ou stéréotypées, avec les actions qu'elles guident (Rizzolatti & Craighero, 2004).

L'implication conjointe de ces régions dans les conditions MOTOR et SUPRA suggère que les participants ont été sensibles au but immédiat (intention motrice) ou général (intention superordonnée) de séquences d'actions. Conformément aux hypothèses du réseau hiérarchique de Hamilton & Grafton, un masquage exclusif de activations SUPRA (but général) par les activations MOTOR (but immédiat) révèle un engagement plus marqué

de IPL dans les conditions SUPRA, ainsi que le recrutement du gyrus frontal inférieur gauche (BA 45/48) et médian droit (BA 46/45). L'implication de l'aire de Brodmann 45 en condition SUPRA n'est pas étonnante : cette région est connue pour son rôle dans l'organisation hiérarchique des séquences, verbales ou motrices (Dominey et al., 2003 ; Fazio et al., 2009 ; Fadiga et al., 2009, pour revue). Son recrutement dans les conditions SUPRA semble témoigner du fait que les participants ne se contentent pas de traiter les actes moteurs en isolation, mais les ordonne hiérarchiquement dans une séquence d'action plus générale, dont l'aboutissement est la forme construite par l'agent.

Enfin, un masquage exclusif des activations MOTOR par les activations SUPRA révèle, bilatéralement, l'implication du sulcus temporal supérieur (STS), dans une partie légèrement plus antérieure que celle habituellement retrouvée dans les tâches de reconnaissance intentionnelle, mais dont les coordonnées correspondent aux activations retrouvées dans les tâches d'imitation d'actions dirigées vers un but (Grèzes & Decéty, 2001 ; Filimon et al., 2007 ; Gazzola et al., 2007b). En général, la partie antérieure du STS code pour l'identification du mouvement *per se* (qu'il soit intentionnel ou non, voir Allison et al., 2000 pour revue). Son activation différentielle en condition MOTOR semble faire écho à la confiance accrue que les participants, dans cette condition, accorde aux informations de cinématique visuelle (contre leurs priors). En miroir, l'absence du STS en condition SUPRA pourrait rendre compte de la tendance qu'ont les participants à négliger l'information de cinématique visuelle et à répondre préférentiellement en direction du modèle *a priori* qu'ils ont formé de la situation (la forme attendue).

6.5.2 Priors et templates intentionnels

Le réseau de structures cérébrales que nous mentionnons ci-dessus comprend des régions dont on soupçonne leur participation au 'système-miroir' humain (Overwalle & Baetens, 2009). Nous trouvons que ces régions sont recrutées lorsque le participant doit prendre une décision au sujet des buts de l'action observée. En revanche, leurs attentes *a priori* concernant ces buts modulent un ensemble de régions extérieures à ce réseau d'activation « miroir », dont le cortex cingulaire postérieur (PCC), le cortex préfrontal ventral (vPFC) et médian (mPFC). Ces régions ne sont néanmoins pas étrangères aux processus de mentalisation, mais leur implication est en général aspécifique, à la fois en termes de cible intentionnelle (croyance, désir, intention) et de matériel utilisé (verbal, iconographique, cinématique) (voir Overwalle & Baetens, 2009, pour revue).

6.5.2.1 Se préparer à reconnaître : le rôle des priors dans la phase de préparation

De manière particulièrement intéressante, nous retrouvons une modulation de l'activité cérébrale par les *priors* durant la phase qui précède l'observation du mouvement (« Main au repos »). En condition SUPRA, cette modulation a été retrouvée dans les parties postérieure et rostrale du cortex cingulaire, ainsi que dans le gyrus frontal médian supérieur (BA 9).

Une récente méta-analyse suggère que la partie rostrale antérieure du cortex cingulaire jouent un rôle privilégié dans le monitoring du rappel d'informations stockées en mémoire – en particulier lorsque le rappel porte sur des informations concurrentes (*conflict monitoring*, cluster 1, figure 15) –, tandis que sa partie postérieure serait plus généralement impliquée dans le rappel d'informations en mémoire épisodique (Maguire, 2001 ; Spiers & Maguire, 2007) et entretiendrait des connexions bilatérales avec les structures hippocampiques (Beckmann et al., 2009) (cluster 2, figure 15). L'implication de ces deux régions durant la phase de préparation semble indiquer un phénomène de rappel similaire. En condition

SUPRA, la phase de préparation présente en effet une main au repos, disposée devant une ébauche de forme. Il est probable que cette ébauche amorce le rappel de la forme qui, parmi toutes celles qui sont congruentes avec cette ébauche, a la probabilité la plus élevée d'être réalisée.

Cette modulation de l'activité cingulaire et préfrontale suggère que les participants, avant même le déclenchement de l'action, ont des attentes particulières sur le but général de la séquence à venir. Or, il n'est pas impossible que cette modulation précoce contribue, sous la forme d'un « bonus au démarrage », à rehausser l'activité des populations neurales recrutées lors de la décision finale. Ce phénomène de « potentialisation » expliquerait dès lors pourquoi, durant la décision, les temps de réaction diminuent significativement plus vite en condition SUPRA que MOTOR lorsque les *priors* des participants augmentent.

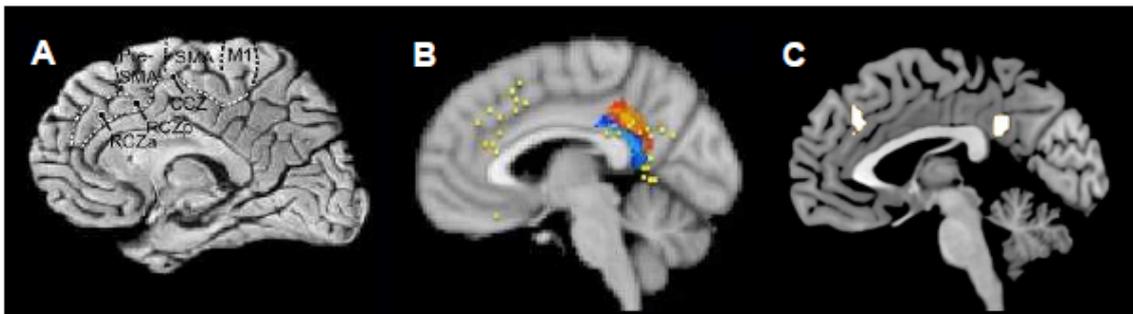


Figure 15. **A.** Parcellisation fonctionnelle du cortex cingulaire (Picard & Strick, 2001). RCZa : gyrus cingulaire rostral antérieur. RCZp : gyrus cingulaire rostral postérieur. CCZ : gyrus cingulaire caudal. **B.** Parcellisation basée sur des études de connectivité (Beckmann et al., 2009) : les points jaunes circonscrivent les clusters 4 (antérieur) et 9 (postérieur) et indiquent les pics d'activation rapportés dans une série de 27 études fonctionnelles. Les activités dans ces deux clusters sont associées au rappel en mémoire épisodique (cluster 4) et/ou à la gestion d'informations conflictuelles stockées en mémoire (cluster 9). **C.** Les régions que nous retrouvons modulées par les priors des participants durant la phase de préparation.

L'implication supplémentaire du gyrus frontal supérieur (BA 9) semble conforter cette hypothèse. Comme nous le suggérons dans la première partie de ce travail (v. p.20), cette région joue un rôle important lorsque le sujet doit sélectionner, parmi plusieurs actions concurrentes, l'action pertinente dans le contexte immédiat de la tâche (Koechlin et al., 2003). Ici, son recrutement semble indiquer que les participants traitent la forme esquissée au centre de la table comme une information contextuelle : l'ébauche de forme *perçue* représente en effet le contexte immédiat dans lequel le sujet doit sélectionner, parmi d'autres actions en compétition, l'action congruente avec la réalisation de la forme *attendue*.

6.5.2.2 Le rôle du vmPFC dans la formation des templates intentionnels

Les conditions dans lesquelles l'action perçue réalise le but attendu devraient naturellement encourager les participants à persévérer en direction de leurs attentes. Cet appariement (*matching*) entre observation et représentation *a priori* pourrait rendre compte de l'effet durable des *priors* sur les performances des sujets. Comme nous le soulignons en introduction, le signal de cet appariement pourrait être pris en charge par le cortex préfrontal ventro-médian (vmPFC). Dans une tâche de décision perceptive, Summerfield et Koechlin ont en effet montré que le gyrus préfrontal ventro-médian (vmPFC) était spécifiquement recruté dans les conditions où l'observation s'ajustait aux attentes préalables des

participants (Summerfield & Koechlin, 2008). Le signal de *matching* subséquent aurait pour effet de « booster » la confiance des participants dans leurs propres *templates* perceptifs.

Or, dans les deux conditions MOTOR et SUPRA, nous observons également une modulation par les *priors* de l'activité située dans la partie médiane du gyrus préfrontal ventral. En l'absence d'analyses de connectivité effective, il nous est pour l'heure difficile de déterminer dans quelle mesure cette représentation des *priors* au niveau ventro-médian exerce une influence sur la hiérarchie corticale en charge du traitement de l'information visuelle. On observe néanmoins que le vmPFC se comporte différemment dans les deux conditions intentionnelles de l'étude : tandis qu'il ne répond qu'aux *priors* des participants en condition SUPRA, son activité est également modulée par la quantité d'information visuelle en condition MOTOR.

Cette distinction nous semble importante, car elle pourrait expliquer la dépendance accrue des participants à leurs *priors* dans la condition SUPRA. Plus largement, elle pourrait constituer une signature cérébrale qui différencie les deux types d'intentions que nous manipulons.

Deux interprétations sont ici possibles. La première nécessite de revenir sur les propriétés individuelles de chaque type d'intentions. En condition MOTOR, l'acte observé (« tourner » par exemple) dénote en effet directement l'intention motrice sous-jacente (« avoir l'intention de tourner le cube ») : dans ce cas, il n'y a rien de plus dans l'intention que dans la kinématique perçue, puisque l'un dénote l'autre sans ambiguïté. Cette relation bi-univoque facilite la comparaison, et le cas échéant, l'appariement entre l'acte perçu et l'intention attendue – ce dont témoigne ici la réponse double du vmPFC aux *priors* et à l'information de kinématique visuelle. Cette hypothèse se trouve par ailleurs confortée par l'implication des régions visuelles associatives (gyrus occipital médian et cuneus) dans la représentation de ces *priors*, mais également de régions prémotrices dont on soupçonne qu'elles possèdent des propriétés « miroir » (pre-SMA, BA 6). En condition MOTOR, la décision du sujet pourrait donc être modélisée comme le résultat d'un processus de comparaison directe entre l'observation et l'*expectation* : lorsqu'il y a *match*, la représentation d'action correspondante est sélectionnée et transmise aux régions motrices via un processus de résonance directe entre les propriétés visuelles de l'acte perçu et la représentation prémotrice attendue.

Or, en condition SUPRA, cette comparaison est compromise par la nature plus abstraite de l'intention considérée : la *forme* que les sujets doivent inférer n'est pas directement lisible dans les actes moteurs qui contribuent à sa réalisation. Le fait que la région préfrontale ventro-médiane ne code que pour les *priors* en SUPRA, et non pour l'information visuelle, pourrait traduire l'impossibilité de cette comparaison entre l'acte perçu et le but général attendu. En l'absence de tout appariement possible, les participants se réfèrent *par défaut* à la forme amorcée durant la phase de préparation. Leur réponse finale ne serait donc plus le résultat d'un processus de résonance directe entre l'observation et l'*expectation*, mais puiserait par défaut dans un répertoire contextuel de réponses congruentes avec la forme qui a été pré-sélectionnée avant même le déclenchement de l'action.

La deuxième interprétation possible n'exclut pas la première : la partie médiane du cortex cingulaire, que l'on retrouve spécifiquement modulé par les *priors* en condition SUPRA, pourrait jouer un rôle central dans le court-circuit (*shortcut*) du processus d'accumulation de preuves visuelles, oblitérant par là-même le processus de comparaison. Ce court-circuit expliquerait l'absence de modulation par l'information visuelle de l'activité préfrontale ventro-médiane, ainsi que la tendance des participants à persévérer dans leurs attentes *a priori* même lorsque l'information visuelle augmente. Cette fonction de court-

circuit est par ailleurs assez cohérente avec les propriétés fonctionnelles de cette région caudale du cortex cingulaire : elle entretient en effet des connections privilégiées avec la partie postérieure du cortex cingulaire (que l'on retrouve modulé par les *priors* durant la phase de préparation), mais également avec les régions pariétales qui codent pour l'information visuelle associée aux actions dirigés vers un but et avec les régions motrices du gyrus précentral (Beckman et al., 2009 ; Picard & Strick, 1996, 2001). Son emplacement dans la hiérarchie cingulaire, au carrefour des régions motrices et pariétales, en fait donc le candidat idéal pour interrompre le flux d'informations perceptives véhiculés depuis les régions pariétales supérieures et biaiser la sélection de la réponse motrice en direction du *template* intentionnel privilégié par le sujet durant la phase de préparation.

6.6 Conclusion préliminaire

Lorsque les *priors* intentionnels portent sur un niveau de représentation de l'action partiellement abstrait des afférences visuomotrices, nous observons ce qui s'apparente à court-circuitage du processus de comparaison entre action perçue et représentation attendue. En l'absence de comparaison, les participants se réfèrent par défaut aux *templates* qu'ils ont préalablement constitués. Ce mode de référence par défaut est à même d'expliquer pourquoi, lorsque le jugement intentionnel porte sur le but superordonné d'une séquence d'actions interchangeable, les participants continuent de se référer à leurs *priors* même lorsque l'information visuelle gagne en fiabilité. Pour être validée, cette hypothèse requiert néanmoins des analyses supplémentaires, en particulier des analyses de connectivité effective entre les régions cérébrales modulées par l'information de cinématique visuelle et les régions qui codent pour les *priors* des participants. Il serait également intéressant de déterminer plus précisément les liens fonctionnels entre ces régions 'priors' et les régions de la hiérarchie décrite par Grafton & Hamilton (2007), que nous retrouvons dans leur quasi-totalité dans la présente étude.

Nous avons choisi ici de ne pas insister sur les conséquences que les résultats – encore préliminaires – de cette étude peuvent avoir pour le débat qui oppose les partisans de la simulation, dans sa version « miroir », aux partisans de la mentalisation pure. Dans l'attente d'analyses supplémentaires, nous rappellerons ces résultats à titre indicatif : en premier lieu, nous observons que les *priors* des participants recrutent des régions extérieures au système-miroir (le mPFC en particulier) et plus habituellement retrouvées dans des tâches impliquant des jugement réflexifs sur des buts, des croyances, ou des problèmes moraux (Van Overwalle & Baetens, 2009, pour revue). Nous observons également que lorsque les participants doivent identifier l'intention sous-jacente à un acte moteur simple (« tourner un cube » par exemple), un processus d'appariement entre observation et *expectation* se met en place, processus qui s'apparente au processus de 'résonance' soutenu par les partisans de l'approche « miroir » de la simulation (Rizzolatti et al., 2001). En revanche, lorsque l'inférence est gouvernée par un but général (la forme construite) qui n'est pas directement lisible dans l'acte moteur observé, ce processus d'appariement est absent et l'on relève une dépendance accrue des participants à leurs attentes *a priori*.

Enfin, nous aimerions souligner le fait que l'implication des cortex préfrontal médian et cingulaire dans la prise en charge des *priors*, ou *templates* intentionnels des participants n'est pas surprenante en soi : comme nous le soulignons plus haut, le mPFC joue un rôle bien connu dans les jugement réflexifs d'actions (de Lange et al., 2008 ; Van der Cruyssen et al., 2009 ; Keysers & Gazzola, 2007). Mais au contraire de la jonction temporo-pariétale, responsable des inférences mentales *transitoires*, le mPFC soutiendrait l'attribution de traits

et qualités *durables* (VanOverwalle, 2009). Or, la nature durable de cette attribution est particulièrement cohérente avec le caractère soutenu, et cumulatif, des influences *top-down* que les *priors* des participants exercent sur leurs propres jugements dans les deux tâches de la présente étude. Nous avons également de bonnes raisons de croire en son implication dans les tâches sociales (DUAL) que nous sommes en train de conduire, dans la mesure où ces tâches induisent des effets réputationnels durables auxquels le mPFC pourrait être particulièrement sensible (Amodio & Frith, 2006).

Reprise & perspectives

L'objectif principal de ce travail était de reconsidérer les processus engagés dans le contrôle et la compréhension de l'action à la lumière de deux modèles hiérarchiques des représentations d'actions. Le second objectif de ce travail visait à déterminer si, dans la schizophrénie, la mise en évidence d'un trouble circonscrit à l'un ou l'autre niveau de la hiérarchie considérée, permettait de mieux rendre compte des spécificités du dysfonctionnement dans la maladie, et s'accompagnait en conséquence d'associations cohérentes avec la sévérité de la symptomatologie.

Nous observons, dans le premier volet de ce travail consacré aux troubles du contrôle de l'action, que les patients schizophrènes présentent des difficultés spécifiques pour le traitement d'informations de nature contextuelle, et que ces difficultés co-varient avec la sévérité des symptômes de désorganisation. Nous avons fait l'hypothèse que cette difficulté pouvait traduire une incapacité plus fondamentale à contrôler de manière soutenue l'organisation hiérarchique du comportement, c'est-à-dire à coordonner les différents segments du comportement avec les buts ou sous-buts qui le gouvernent. Dans une étude complémentaire de connectivité effective, nous montrons par ailleurs que ce trouble contextuel s'accompagne de liens fonctionnels anormaux entre les régions préfrontales en charge du traitement des indices de contexte immédiat et du contrôle de l'épisode temporel en cours. Ces relations dysfonctionnelles pourraient traduire, selon nous, les difficultés qu'ont certains patients à affiner leurs représentations temporelles des situations en y intégrant des indices de contexte pertinents. Ces difficultés feraient par ailleurs écho au caractère souvent « dissonant » des comportements désorganisés de la maladie, dans les registres moteur comme verbal.

Un déficit en matière d'intégration contextuelle peut naturellement conduire à une forme de « décontextualisation » de l'action susceptible d'impacter à la fois sur l'organisation des comportements en première personne, mais également sur la capacité à planifier ses actions dans le contexte des plans, ou des buts, d'autrui. Le second volet de ce travail s'attachait précisément à explorer la capacité réduite des patients schizophrènes à juger ou se représenter les intentions d'un tiers. Ces difficultés d'attribution et de représentation mentales ont fait l'objet d'investigations détaillées, mais souvent contradictoires, et nous avons proposé de les explorer à un niveau d'analyse plus fin du fonctionnement métacognitif.

Pour ce faire, nous avons pointé l'importance de considérer le caractère composite de la notion d' « intention », en insistant sur la variété des relations (*one-to-one*, *one-to-many*) qu'une action peut entretenir avec le ou les buts qu'elle poursuit. Nous avons fait l'hypothèse que la variété de ces relations devait s'exprimer sous la forme d'une interaction variable des processus les plus élémentaires en jeu dans l'inférence intentionnelle. Reconnaître une intention suppose en effet de traiter *au moins* deux types d'information distincts : les informations visuelles de kinématique (véhiculées par le comportement observé), et les informations, ou attentes préalables que l'observateur formule à l'égard du comportement dont il cherche à découvrir le but. Dans une série d'études comportementales, nous avons montré que ces deux types d'information interagissaient différemment selon le type d'intention considéré, c'est-à-dire i) selon la relation que l'action observée entretient avec

l'intention qui la cause (intention motrice vs. superordonnée) ou ii) selon que la situation d'action induit, ou non, des attentes domaine-spécifiques (intention sociale vs. non-sociale). Nous avons proposé d'interpréter ces résultats à la lumière d'un modèle hiérarchique de compréhension de l'action dans le cadre duquel l'inférence intentionnelle est modélisée comme le produit d'une cascade d'influences *top-down*, générées et intégrées à chaque niveau de la hiérarchie, et remises à jour à chaque nouvelle observation (v. figure 16, p.247). Nous suggérons en outre que cette remise à jour pourrait s'effectuer selon une dynamique formalisée par une généralisation du théorème de Bayes (modèle log(proba)), optimisé pour chaque individu (horizon mnésique).

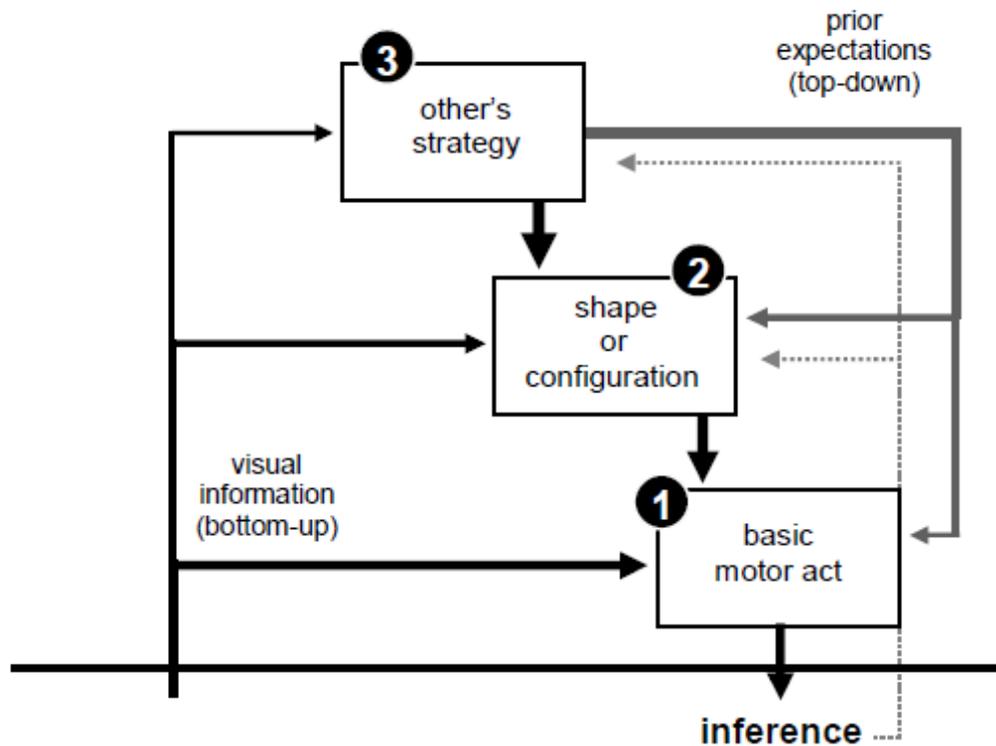


Figure 16. Modèle hiérarchique des représentations d'actions intentionnelles. L'interaction entre information visuelle et connaissance a priori varie selon le niveau intentionnel de l'action considérée (moteur, superordonné, social). Cette variation dépend de la relation que l'action observée entretient avec l'intention qui la cause : (1) lorsque cette relation est de type univoque (one-to-one), l'inférence table principalement sur l'information perceptive, véhiculée par la cinématique du mouvement ; (2) lorsque cette relation est équivoque (one-to-many), l'inférence s'affranchit progressivement des afférences sensorielles pour privilégier les connaissances a priori relatives aux causes les plus probables du mouvement observé. Des intentions motrices aux intentions sociales (3), on observe une influence accrue de ces attentes a priori sur la sélection de l'action exécutée (1), ou de la configuration arrangée (2) par les agents de l'interaction. Dans ce modèle, l'inférence intentionnelle est le produit d'une cascade d'influences topdown, produites et intégrées à chaque niveau de la hiérarchie. A chaque essai, la réponse du sujet met à jour la probabilité (lignes en pointillés) associée à chaque occurrence d'actions, formes, configurations, ou stratégies.

Conformément à notre hypothèse initiale, cette série de quatre études comportementales a été répliquée dans la schizophrénie, chez 4 groupes de patients

appariés. Deux profils distincts de mentalisation émergent. Nous observons, d'une part, que les patients à symptomatologie productive dominante (hallucinations, délires, troubles de la pensée formelle) tendent à accorder un crédit excessif à leurs *attentes* préalables. En revanche, la sévérité de la symptomatologie négative (retrait social, émoussement affectif, pauvreté du langage et de la pensée) co-varie avec une tendance exagérée à se concentrer sur l'information de kinématique visuelle, véhiculée par les mouvements de l'agent observé. Nous avons fait l'hypothèse que l'hétérogénéité des performances rapportées dans la littérature pouvait être redevable de ce pattern de dépendance spécifique – aux informations *a priori* ou aux informations perceptives de kinématique –, variable selon la dimension clinique du patient et accentué par les caractéristiques de l'intention considérée.

Enfin, la dernière étude de ce travail s'inscrivait dans la continuité des hypothèses que nous avons formulées initialement : la mise en évidence d'une dépendance normale (sujets sains), ou excessive (patients schizophrènes), aux informations perceptives ou *a priori* devait pouvoir se traduire par l'engagement d'un réseau de structures cérébrales propre à chaque classe d'intention (selon, précisément, le type d'information qu'elle promeut). Les résultats préliminaires de l'étude conduite en neuroimagerie démontrent que les régions impliquées dans le jugement d'intentions sont, d'une part, sensibles aux deux types d'information manipulés durant la tâche, et, d'autre part, que ces informations contribuent différemment à l'inférence d'intentions motrices ou superordonnées. En condition superordonnée, nous observons un pré-engagement des attentes du participant avant l'observation du mouvement lui-même, et une référence *par défaut* à ce *template* intentionnel préalablement constitué – ce dont témoigne l'absence de modulation par l'information visuelle de l'activité des régions qui codent pour la représentation de ce *template*. Ces résultats, confinés aux conditions non-sociales, sont toutefois encore préliminaires et nécessiteraient la conduite d'analyses supplémentaires, ainsi qu'une comparaison directe avec les conditions sociales développées dans l'étude comportementale à l'origine de ce travail.

Considérés ensemble, ces résultats présentent une cohérence qui appelle, selon nous, une ligne de recherche prometteuse. Nous devons toutefois souligner que la position théorique qui fait l'originalité de ce travail, pourrait également en constituer l'une des principales limites. Nous avons en effet formulé l'hypothèse que le jugement d'intentions mobilisait des processus dont le fonctionnement s'ajustait au cadre plus général du formalisme Bayésien. Ce formalisme a démontré son efficacité dans des champs aussi variés que l'apprentissage, la perception ou le raisonnement causal. Plus généralement, il est parfaitement adapté pour rendre compte de la manière dont un observateur parvient, en présence de signaux tronqués ou ambigus, à inférer avec succès les variables cachées de son environnement. En inscrivant notre travail dans ce cadre, nous avons donc choisi d'aborder le problème de la lecture intentionnelle en mobilisant les concepts et les outils d'un formalisme dont le champ d'application excède, précisément, celui de la mentalisation. Cet acte délibéré répondait à la nécessité de décomposer la fonction étudiée en processus plus élémentaires, susceptibles dans la pathologie de dysfonctionnements localisés ; il n'en reste pas moins que le formalisme Bayésien reste ici limité à l'observation d'actions réalisées en troisième personne, et ne nous dit rien des mécanismes en jeu dans la représentation d'actions intentionnelles *en première personne*. Or, la question de la représentation et de l'implémentation de ses propres buts pourrait engager des motivations hors de portée du formalisme adopté dans ce travail. L'initiation et le suivi de ma propre action, par exemple, est susceptible de reposer sur des attentes préalables codées sous la forme de *valeurs internes*, ou de *raisons*, dont seule une étude de l'action en première personne est à même

de saisir l'importance, en particulier pour la compréhension de ce qui fait la spécificité des comportements autonomes humains.

Références bibliographiques

– Partie I – L'architecture du contrôle cognitif dans la schizophrénie

- Allan, P., Amaro, E., Fu, C.H., Williams, S.C., Brammer, M.J., Johns, L.C., & McGuire, P.K. (2007). Neural correlates of the misattribution of speech in schizophrenia. *Br J Psychiatry*, 190: 162-9.
- Anderson, S.W., Damasio, H., Jones, R.D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *J Clin Exp Neuropsychol*, 13, 909-22.
- Andreasen NC. The Scale for the Assessment of Negative Symptoms (SANS). Iowa City, The University of Iowa, 1983.
- Andreasen NC. The Scale for the Assessment of Positive Symptoms (SAPS). Iowa City, The University of Iowa, 1984.
- Andreasen, N.C., Rezaï, K., Alliger, R., Swayze, V.W., Flaum, M., Kirchner, P., Cohen, G., & O'Leary, D.S. (1992). Hypofrontality in neuroleptic-naïve patients and in patients with chronic schizophrenia. Assessment with xenon 133 single-photon emission computed tomography and the Tower of London. *Arch Gen Psychiatry*, 49, 943-958.
- Asaad WF, Rainer G, Miller EK. Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol*. 2000; 84(1):451-459.
- Axelrod BN, Goldman RS, Heaton RK, Curtiss G, Thompson LL, Chelune GJ, Kay GG. Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *J Clin Exp Neuropsychol* 1996; 18:338-42.
- Baddeley, A.D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed), *Recent advances in learning and motivation*, vol. 8. New York : Academic Press.
- Baddeley, A.D. (2000b). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences* 4, 417–423.
- Badre D, D'Esposito M. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J Cogn Neurosci*. 2007; 19(12):2082-2099.
- Barch DM, Carter CS, Braver TS, Sabb FW, MacDonald A3, Noll DC, Cohen JD. Selective deficits in prefrontal cortex function in medication-naïve patients with schizophrenia. *Arch Gen Psychiatry*. 2001; 58(3):280-288.
- Barch DM, Csernansky JG, Conturo T, Snyder AZ. Working and long-term memory deficits in schizophrenia: is there a common prefrontal mechanism? *J Abnorm Psychol*. 2002; 111(3):478-494.
- Barch DM, Carter CS, MacDonald AW3, Braver TS, Cohen JD. Context-processing deficits in schizophrenia: diagnostic specificity, 4-week course, and relationships to clinical symptoms. *J Abnorm Psychol*. 2003; 112(1):132-143.

- Barch DM. The cognitive neuroscience of schizophrenia. *Annu Rev Clin Psychol.* 2005; 1:321-353.
- Barch DM, Csernansky JG. Abnormal parietal cortex activation during working memory in schizophrenia: verbal phonological coding disturbances versus domain-general executive dysfunction. *Am J Psychiatry.* 2007; 164(7):1090-1098.
- Berlyne, D. (1957). Uncertainty and conflict: a point of contact between information-theory and behavior-theory concepts. *Psychological Review*, 64, 329-339.
- Bleuler, E. (1911). *Dementia Praecox ou Groupe des schizophrénies*. Préface de Viallard, A., Paris, EPEL, GREC, 1993.
- Bozikas VP, Kosmidis MH, Kiosseoglou G, Karavatos A (2006). Neuropsychological profile of cognitively impaired patients with schizophrenia. *Compr Psychiatry* 47: 136–143.
- Braff, D.L. (1992). Information processing and attention dysfunctions in schizophrenia. *Schizophrenia Bulletin*, 19,233-259.
- Braver TS, Barch DM, Cohen JD. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol Psychiatry.* 1999; 46(3):312-328.
- Braver TS, Cohen JD. Dopamine, cognitive control, and schizophrenia: the gating model. *Prog Brain Res* 1999;121:327-349.
- Braver TS, Reynolds JR, Donaldson DI. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron.* 2003; 39(4):713-726.
- Burgess, P.W., & Shallice, T. (1996a). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, 34, 263-272.
- Burgess, P.W., & Shallice, T. (1996b). Bizarre responses, rule detection and frontal lobe lesions. *Cortex*, 32, 241-259.
- Burglen, F., Marczewski, P., Mitchell, K.J., Van der Linden, M., Johnson, M.K., Danion, J-M., & Salamé, P. (2004). Impaired performance in a working memory binding task in patients with schizophrenia. *Psychiatry Research*, 125, 247-255.
- Callicott JH, Bertolino A, Mattay VS, Langheim FJ, Duyn J, Coppola R, Goldberg TE, Weinberger DR. Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. *Cereb Cortex.* 2000; 10(11):1078-1092.
- Callicott JH, Mattay VS, Verchinski BA, Marenco S, Egan MF, Weinberger DR. Complexity of prefrontal cortical dysfunction in schizophrenia: more than up or down. *Am J Psychiatry.* 2003; 160(12):2209-2215.
- Cannon TD, Glahn DC, Kim J, Van Erp TGM, Karlsgodt K, Cohen MS, Nuechterlein KH, Bava S, Shirinyan D. Dorsolateral prefrontal cortex activity during maintenance and manipulation of information in working memory in patients with schizophrenia. *Arch Gen Psychiatry.* 2005; 62(10):1071-1080.
- Carter, C.S., Mintun, M., Nichols, T., & Cohen, J.D. (1997). Anterior cingulate gyrus dysfunction and selective attention deficits in schizophrenia: [15O]H₂O PET study during single-trial Stroop task performance. *American Journal of Psychiatry*, 154, 1670-1675.

- Carter CS, MacDonald AW3, Ross LL, Stenger VA. Anterior cingulate cortex activity and impaired self-monitoring of performance in patients with schizophrenia: an event-related fMRI study. *Am J Psychiatry*. 2001; 158(9):1423-1428.
- Chambon V, Franck N, Koechlin E, Fakra E, Ciuperca G, Azorin J, Farrer C. The architecture of cognitive control in schizophrenia. *Brain*. 2008; 131:962-970.
- Chapman LJ, Chapman JP. Problems in the measurement of cognitive deficit. *Psychological Bulletin* 1973b; 79:380-385.
- Chapman LJ, Chapman JP. Commentary on two articles concerning generalized and specific cognitive deficits. *J Abnorm Psychol* 2001; 110:31-9. Review.
- Cohen JD, Servan-Schreiber D. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychol Rev*. 1992; 99(1):45-77.
- Cohen JD, Barch DM, Carter C, Servan-Schreiber D. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. *J Abnorm Psychol* 1999; 108:120-33.
- Collette, F., Van der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Hum Brain Mapp*, 158, 524-547.
- Condray, R. (2005). Language disorder in schizophrenia as a developmental learning disorder. *Schizophrenia Research*, 73, 5-20. Review.
- Dalrymple-Alford, J.C., Kalders, A.S., Jones, R.D., & Watson, R.W. (1994). A central executive deficit in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, Vol. 57, 360-367.
- Danion JM, Rizzo L, Bruant A. Functional mechanisms underlying impaired recognition memory and conscious awareness in patients with schizophrenia. *Arch Gen Psychiatry*. 1999; 56(7):639-644.
- Diedrichsen J, Shadmehr R. Detecting and adjusting for artifacts in fMRI time series data. *Neuroimage*. 2005; 27(3):624-634.
- Dominey P., Hoen M, Blanc JM, Lelekov-Boissard T. Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. *Brain and Language* 2003; 86:207-225.
- Eisenberg, D.P. & Berman, K.F. (2009). Executive Function, Neural Circuitry, and Genetic Mechanisms in Schizophrenia. *Neuropsychopharmacology*, in press.
- Elvevag B, Duncan J, McKenna P. The use of cognitive context in schizophrenia: An investigation. *Psychological Medicine* 2000; 30:885-897
- Evans AC, Kamber M, Collins DL, MacDonald D. A mri-based probabilistic atlas of neuroanatomy. In: Shorvon S, Fish D, Andermann, Bydder GM, Stefan H (eds). *Magnetic Resonance Scanning And Epilepsy*. NATO ASI series A: New York Plenum; 1994: 263-274.
- Friston KJ, Holmes AP, Worsley KJ, Poline J-B, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*. 1994; 2(4):189-210.

- Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. *Human Brain Mapping*. 1995; 3(3):165-189.
- Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ. The trouble with cognitive subtraction. *Neuroimage*. 1996; 4(2):97-104.
- Frith, C.D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove : Lawrence Erlbaum Associates.
- Fuller, R., & Jahanshahi, M.(1999). Impairment of willed actions and use of advance information for movement preparation in schizophrenia. *J Neurol Neurosurg Psychiatry*, 66, 502-509.
- Funahashi S. Neuronal mechanisms of executive control by the prefrontal cortex. *Neuroscience research* 2001; 39:147-165. Review.
- Fuster, J.M. (1997). *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe*, third ed., Lippincott-Raven, Philadelphia.
- Fuster, J.M. (2004). Upper processing stages of the perception-action cycle. *Trends in Cognitive Science* 8,143-145.
- Glahn D, Ragland J, Abramoff A, Barrett J, Laird AR, Bearden CE, Velligan DI. Beyond hypofrontality: a quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Hum Brain Mapp*. 2005; 25:60-69.
- Godefroy O, Cabaret M, Petit-Chenal V, Pruvo JP, Rousseaux M. Control functions of the frontal lobes. Modularity of the central-supervisory system? *Cortex* 1999; 35:1-20.
- Goldman, R.S., Axelrod, B.N., & Tompkins, L.M. (1992). Effect of instructional cues on schizophrenic patients' performance on the Wisconsin Card Sorting Test. *American Journal of Psychiatry*, 149, 1718-1722.
- Goldman-Rakic PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Plum F, Mountcastle V, eds. *Handbook Of Physiology: The Nervous System V*. Bethesda, MD: American Physiological Society; 1987: 373–417.
- Goldman-Rakic, P.S. (1994). Working memory dysfunction in schizophrenia. *J Neuropsychiatry Clin Neurosci*, 6, 348-357. Review.
- Goldman-Rakic PS. The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philos Trans R Soc Lond B Biol Sci*. 1996; 351:1445-1453.
- Goldman-Rakic, P.S., & Selemon, L.D. (1997). Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophrenia Bulletin*, 23, 437-458. Review.
- Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigle-type card sorting problem. *Journal of Experimental Psychology*, 32, 404-411.
- Grapperon, J., & Delage, M. (1999). Stroop test and schizophrenia. *Encephale*, 25, 50-58.
- Gras-Vincendon, A., Danion, J.M., Grangé, D., Bilik, M., Willard-Schroeder, D., Sichel, J.P., & Singer, L. (1994). Explicit Memory, repetition priming and cognitive skill learning in schizophrenia. *Schizophrenia Research*, 13, 117-126.

- Green, M.F., & Nuechterlein, K.H., (1999). Should schizophrenia be treated as a neurocognitive disorder? *Schizophrenia Bulletin*, 25, 309-319.
- Hartman, M., Seketee M.C., Silva, S., Lanning, K., & Andersson, C. (2005). Wisconsin Card Sorting Test performance in schizophrenia : the role of working memory. *Schizophr Res*, 63, 201-17.
- Harvey , I., Ron, M. A., Du Boulay, G., Wicks, D., Lewis, S. W. & Murray, R. M. (1993). *Reduction of cortical volume in schizophrenia on magnetic resonance imaging. Psychological Medicine*, 23: 591-604.
- Hazy TE, Frank MJ, O'reilly RC. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos Trans R Soc Lond B Biol Sci*. 2007; 362(1485):1601-1613.
- Heckers S, Rauch SL, Goff D, Savage CR, Schacter DL, Fischman AJ, Alpert NM. Impaired recruitment of the hippocampus during conscious recollection in schizophrenia. *Nat Neurosci*. 1998; 1(4):318-323.
- Heckers S, Weiss AP, Deckersbach T, Goff DC, Morecraft RJ, Bush G. Anterior cingulate cortex activation during cognitive interference in schizophrenia. *Am J Psychiatry*. 2004; 161(4):707-715.
- Heinrichs, R.W., & Zakzanis, K.K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*, 12, 426-445.
- Hemsley DR, Phil M. The Schizophrenic Experience: Taken Out of Context? *Schizophrenia Bulletin* 2005; 31:43-53
- Henik A, Salo R. Schizophrenia and the stroop effect. *Behav Cogn Neurosci Rev*. 2004; 3(1):42-59.
- Hepp, H.H., Maier, S., Hermle, L., & Spitzer, M. (1996). The Stroop effect in schizophrenic patients. *Schizophrenia Research*, 22, 187-195.
- Hill, S.K., Ragland, J.D., Gur, R.C., & Gur, R.E. (2001). Neuropsychological differences among empirically derived clinical subtypes of schizophrenia. *Neuropsychology*, 15, 492-501.
- Hofer A, Weiss EM, Golaszewski SM, Siedentopf CM, Brinkhoff C, Kremser C, Felber S, Fleischhacker WW. An fmri study of episodic encoding and recognition of words in patients with schizophrenia in remission. *Am J Psychiatry*. 2003; 160(5):911-918.
- Hofer A, Weiss EM, Golaszewski SM, Siedentopf CM, Brinkhoff C, Kremser C, Felber S, Fleischhacker WW. Neural correlates of episodic encoding and recognition of words in unmedicated patients during an acute episode of schizophrenia: a functional mri study. *Am J Psychiatry*. 2003; 160(10):1802-1808.
- Hoff, A.L., Riordan, H., O'Donnell, D.W., Morris, L., & DeLisi, L.E. (1992). Neuropsychological functioning of first-episode schizophreniform patients. *American Journal of Psychiatry*, 149, 898-903.
- Jones, S.H., Hemsley, D.R., & Gray, J.A. (1991). Contextual effects on choice reaction time and accuracy in acute and chronic schizophrenics. Impairment in selective attention or in the influence of prior learning? *British Journal of Psychiatry*, 159, 415-21.

- Kerns JG, Berenbaum H. The relationship between formal thought disorder and executive functioning component processes. *J Abnorm Psychol* 2003; 112:339-52.
- Kerns JG, Cohen JD, MacDonald AW3, Johnson MK, Stenger VA, Aizenstein H, Carter CS. Decreased conflict- and error-related activity in the anterior cingulate cortex in subjects with schizophrenia. *Am J Psychiatry*. 2005; 162(10):1833-1839.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302: 1181-1184.
- Koechlin E, Jubault T. Broca's area and the hierarchical organization of human behavior. *Neuron*. 2006; 50(6):963-974.
- Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci*. 2007; 11(6):229-235.
- Kraepelin, E. (1919). *Dementia Praecox and the Paraphrenia*. Edinburg: E. & S. Livingston.
- Kravariti E, Dixon T, Frith C, Murray R, McGuire P. Association of symptoms and executive function in schizophrenia and bipolar disorder. *Schizophr Res* 2005; 74:221-31.
- Kurtz MM, Moberg PJ, Ragland JD, Gur RC, Gur RE (2005). Symptoms versus neurocognitive test performance as predictors of psychosocial status in schizophrenia: a 1- and 4-year prospective study. *Schizophr Bull* 31: 167–174.
- Lezak MD. Newer Contributions to the Neuropsychological Assessment of Executive Functions. *Journal of Head Trauma Rehabilitation*, 1993; 8:24-31.
- Lysaker, P., Bell, M., & Beam-Goulet, J. (1997). Wisconsin Card Sorting Test and work performance in schizophrenia. *Psychiatry Research*, 56, 45-51
- MacDonald AW3, Carter CS. Event-related fmri study of context processing in dorsolateral prefrontal cortex of patients with schizophrenia. *J Abnorm Psychol*. 2003; 112(4):689-697.
- MacDonald AW3, Carter CS, Kerns JG, Ursu S, Barch DM, Holmes AJ, Stenger VA, Cohen JD. Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *Am J Psychiatry*. 2005; 162(3):475-484.
- Mahurin, R.K., Velligan, D.I., & Miller, A.L. (1998). Executive-frontal lobe cognitive dysfunction in schizophrenia: a symptom subtype analysis. *Psychiatry Research*, 79, 139-49.
- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. *Neuroimage*. 2003; 19(3):1233-1239.
- Mamah, D., Wang, L., Barch, D, de Erausquin, G.A., Gado, M., & Csernansky, J.G. (2007). Structural analysis of the basal ganglia in schizophrenia. *Schizophrenia Res*, 89: 59-71.
- Marczewski, P., Van der Linden, M., & Laroï, F. (2001). Further investigation of the Supervisory Attentional System in schizophrenia : Planning, inhibition, and rule abstraction. *Cognitive Neuropsychiatry*, 6, 175-192.

- Mattson, A.J., & Levin, H.S. (1990). Frontal lobe dysfunction following closed head injury. A review of the literature. *J Nerv Ment Dis*, 178, 282-291. Review.
- Mechelli, A., Allen, P., Amaro, E. Jr., Fu, C.H., Williams, S.C., Brammer, M.J., Johns, L.C., & McGuire, P.K. (2007). Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. *Hum Brain Mapp*, 28: 1213-22.
- Meulemans, T. (2004). *Neuropsychologie des fonctions exécutives*, coll. « Neuropsychologie », Solal Editeurs.
- Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*. 2001; 24:167-202.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, 41:49-100.
- Morice, R., & Delahunty, A. (1996). Frontal/executive impairments in schizophrenia. *Schizophr Bulletin*, 22, 125-137.
- Mountain, M.A., & Snow, W.G. (1993). Wisconsin Card Sorting Test as a measure of frontal pathology : a review. *Clinical Psychologist*, 9, 90-100.
- Mortimer, A.M. (1996). Cognitive function in schizophrenia- Do neuroleptics make a difference? *Pharmacol Biochem Behav*, 56, 789-795.
- Musso M, Moro A, Glauche V, Rijntjes M, Reichenbach J, Buchel C, & Weiller, C. Broca's area and the language instinct. *Nature Neuroscience* 2003; 6:774–781.
- Nuechterlein, K.H. (1977). Reaction time and attention in schizophrenia: A critical evaluation of the data and theories. *Schizophrenia Bulletin*, 3, 373-428.
- O'Leary, D.S., Flaum, M., Kesler, M.L., Flashman, L.A., Arndt, S., & Andreasen, N. (2000). Cognitive correlates of the negative, disorganized, and psychotic symptom dimensions of schizophrenia. *J Neuropsychiatry Clin Neurosci*, 12, 4-15.
- Park, S., Lee, J., Folley, B., & Kim, J. (2003). Schizophrenia: Putting context in context. *Behavioral and Brain Sciences*, 26, 98-99.
- Palmer, B.W., & Heaton, R.K. (2000). *Cognition in schizophrenia : "Impairments, Importance and Treatment Strategies"*, Oxford University Press, New York.
- Passingham, R.E. (1993). *The Frontal Lobes and Voluntary Action*. Oxford University Press, New York.
- Perlstein, W.M., Carter, C.S., Barch, D.M., & Baird, J.W. (1998). The Stroop task and attention deficits in schizophrenia: a critical evaluation of card and single-trial Stroop methodologies. *Neuropsychology*, 12, 414-425.
- Perlstein WM, Carter CS, Noll DC, Cohen JD. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am J Psychiatry*. 2001; 158(7):1105-1113.
- Perlstein WM, Dixit NK, Carter CS, Noll DC, Cohen JD. Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biol Psychiatry*. 2003; 53(1):25-38.

- Petrides M, Alivisatos B, Evans AC, Meyer E. Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. *Proc Natl Acad Sci U S A*. 1993; 90(3):873-877.
- Petrides M. Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philos Trans R Soc Lond B Biol Sci*. 1996; 351(1346):1455-61; discussion 1461-2.
- Peuskens J, Demily C, Thibaut F. Treatment of cognitive dysfunction in schizophrenia. *Clin Ther* 2005; 27 Suppl A: S25-37.
- Posada, A., & Franck, N. (2002). Use and automation of a rule in schizophrenia. *Psychiatry Research*, 109,289-296.
- Purdon, S.E., Jones, B.D., Stip, E., Labelle, A., Addington, D., David, S.R., Breier, A., & Tollefson, G.D. (2000). Neuropsychological change in early phase schizophrenia during 12 months of treatment with olanzapine, risperidone, or haloperidol. *Archives of General Psychiatry*, 57, 249–258.
- Rabbitt, P. (1997). *Methodology of frontal and executive function*. Hove: East Sussex, Psychology Press.
- Ragland, J.D. (2003). Profiles of neuropsychologic function in schizophrenia. *Curr Psychiatry Rep*, 5, 299-302. Review.
- Ramsey NF, Koning HAM, Welles P, Cahn W, van der Linden JA, Kahn RS. Excessive recruitment of neural systems subserving logical reasoning in schizophrenia. *Brain*. 2002; 125(Pt 8):1793-1807.
- Rankin, P.M., & O'Carroll, P.J. (1995). Reality discrimination, reality monitoring and disposition towards hallucination. *British Journal of Clinical Psychology*, 34, 517–528.
- Rizzo, L., Danion, J.-M., Van der Linden, M., Grange, D., & Rohmer, J.G. (1996). Impairment of memory for spatial context in schizophrenia. *Neuropsychology*, 10, 376–384.
- Rizzo L, Danion JM, van der Linden M, Grange D. Patients with schizophrenia remember that an event has occurred, but not when. *Br J Psychiatry*. 1996; 168(4):427-431.
- Sakai K, Passingham RE. Prefrontal interactions reflect future task operations. *Nat Neurosci*. 2003; 6(1):75-81.
- Saykin, AJ, Gur, RC, & Gur, RE (1991). Neuropsychological function in schizophrenia, selective impairment in learning and memory. *Arch Gen Psychiatry*, 48:618-623.
- Schooler, C., Roberts, B., & Cohen, R. (2008). Context, complexity, and cognitive processing in schizophrenia. *Cogn Neuropsychiatry*, 13: 250-66.
- Servan-Schreiber D, Cohen JD, Steingard S. Schizophrenic deficits in the processing of context. a test of a theoretical model. *Arch Gen Psychiatry*. 1996; 53(12):1105-1112.
- Shallice, T. (1982). Specific impairments of planning. *Philos Trans R Soc Lond B Biol Sci*, 298, 199-209.
- Shallice, T., (1988). *From Neuropsychology to Mental Structure*. Cambridge Univ. Press, Cambridge.

- Shallice, T., Burgess, P.W., & Frith, C.D. (1991). Can the neuropsychological case-study approach be applied to schizophrenia? *Psychol Medicine*, 21, 661-673.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423, 623, 656.
- Snitz BE, MacDonald A3, Cohen JD, Cho RY, Becker T, Carter CS. Lateral and medial hypofrontality in first-episode schizophrenia: functional activity in a medication-naive state and effects of short-term atypical antipsychotic treatment. *Am J Psychiatry*. 2005; 162(12):2322-2329.
- Spitzer, M. (1993). The psychopathology, neuropsychology, and neurobiology of associative and working memory in schizophrenia. *Eur Arch Psychiatry Clin Neurosci*, 243, 57-70.
- Stevens AA, Goldman-Rakic PS, Gore JC, Fulbright RK, Wexler BE. Cortical dysfunction in schizophrenia during auditory word and tone working memory demonstrated by functional magnetic resonance imaging. *Arch Gen Psychiatry* 1998; 55:1097-103
- Stratta, P., Daneluzzo, E., Prosperini, P., Bustini, M., Mattei, P., & Rossi, A. (1997). Is Wisconsin Card Sorting Test performance related to 'working memory' capacity? *Schizophrenia Research*, 27, 11-19.
- Stratta P, Daneluzzo E, Bustini M, Casacchia M, Rossi A. Schizophrenic deficit in the processing of context. *Arch Gen Psychiatry* 1998; 55:186–188.
- Stratta P, Daneluzzo E, Bustini M, Prosperini P, Rossi A. Processing of context information in schizophrenia: relation to clinical symptoms and wcst performance. *Schizophr Res*. 2000; 44(1):57-67.
- Stuss, D.T., & Benson D.F. (1986). *The Frontal Lobes*. Raven Press, New York.
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Talairach J TP. *Co-Planar Stereotaxic Atlas Of The Human Brain*. New York, NY: Thieme Medical Publishers; 1988.
- Tan H, Choo W, Fones CSL, Chee MWL. Fmri study of maintenance and manipulation processes within working memory in first-episode schizophrenia. *Am J Psychiatry*. 2005; 162(10):1849-1858.
- Tan H, Sust S, Buckholtz JW, Mattay VS, Meyer-Lindenberg A, Egan MF, Weinberger DR, Callicott JH. Dysfunctional prefrontal regional specialization and compensation in schizophrenia. *Am J Psychiatry*. 2006; 163(11):1969-1977.
- Tulving E. Episodic memory: from mind to brain. *Annu Rev Psychol*. 2002; 53:1-25.
- Van der Does, A.J.W, & Van den Bosch, R.J. (1992). What determines Wisconsin Card Sorting Test performance in schizophrenia? *Clinical Psychology Review*, 12, 567-583.
- Van der Linden, M., Meulemans, T., Seron, X., Coyette, F., Andrès, P., & Prairial, C. (2000). L'évaluation des fonctions exécutives. In X. Seron & M. Van der Linden (Eds.), *Traité de Neuropsychologie Clinique, Tome 1* (pp. 275-300). Marseille: Solal.
- Velakoulis D, Wood SJ, Wong MT, McGorry PD, Yung A, Phillips L, Smith D, Brewer W, Proffitt T, Desmond P, & Pantelis C. (2006). Hippocampal and amygdala volumes

- according to psychosis stage and diagnosis: a magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals. *Arch Gen Psychiatry*, 63: 139-49.
- Velligan, D.I., & Bow-Thomas, C.C. (1999). Executive function in schizophrenia. *Semin Clin Neuropsychiatry*, 4,24-33. Review.
- Volz, H.P., Gaser, C., Hager, F., Rzanny, R., Mentzel, H.J., Kreitschmann-Andermahr, I., Kaiser, W.A., & Sauer, H. (1997). Brain activation during cognitive stimulation with the Wisconsin Card Sorting Test--a functional MRI study on healthy volunteers and schizophrenics. *Psychiatry Research*, 75, 145-157.
- Wang, L., Mamah, D., Harms, M.P., Karnik, M., Price, J.L., Gado, M.H., Thompson, P.A., Barch, D.M., Miller, M.I., & Csernansky, J.G. (2008). Progressive deformation of deep brain nuclei and hippocampal-amygdala formation in schizophrenia. *Biol J Psychiatry*, 64: 1060-8.
- Wechsler D. Wechsler Adult intelligence Scale (3rd ed.) Administration and Scoring Manual. Psychological Corporation/Harcourt Brace, San Antonio (TX); 1997.
- Weinberger, D.R., & Berman, K.F. (1996). Prefrontal function in schizophrenia: confounds and controversies. *Philos Trans R Soc Lond B Biol Sci*, 351, 1495-1503. Review.
- Weiss AP, Schacter DL, Goff DC, Rauch SL, Alpert NM, Fischman AJ, Heckers S. Impaired hippocampal recruitment during normal modulation of memory performance in schizophrenia. *Biol Psychiatry*. 2003; 53(1):48-55.
- Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. *J Clin Psychiatry*. 2003; 64(6):663-667.
- Zalla, T., Joyce, C., Szoke, A, Schurhoff, F., Pillon, B., Komano, O., Perez-Diaz, F., Bellivier, F., Alter, C., Dubois, B., Rouillon, F., Houde, O., & Leboyer, M. (2004). Executive dysfunctions as potential markers of familial vulnerability to bipolar disorder and schizophrenia. *Psychiatry Research*, 12, 207– 217.

– Partie II – Les mécanismes d'attribution d'intentions chez le sujet sain et le patient schizophrène

- Abu-Akel, A., & Bailey, A.L. (2000). The possibility of different forms of theory of mind, *Psychological Medicine*, 30, 735-738.
- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, 11, 1109-16.
- Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 4, 267–278.
- Amodio, D.M. & Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 7, 268-77. Review.

- André, J.B., & Day, T. (2007). Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. *J Theor Biol*, 247, 11-22.
- Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition*. Princeton, Princeton University Press.
- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2006). Bayesian Models of Human Action Understanding. In Y. Weiss, B. Scholkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18, pp. 99-106), MIT Press.
- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2007). Goal Inference as Inverse Planning, In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*.
- Barbalat G., Chambon V., Franck N., Koechlin E., Farrer C. (2009). Organization of cognitive control within lateral prefrontal cortex in schizophrenia, *Archives of General Psychiatry*, 66, 1-10.
- Barnes, C.L. & Pandya, D.N. (1992). Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 318: 222-44.
- Baron-Cohen, S. *et al.* (1985). Does the autistic child have a 'theory of mind' ? *Cognition*, 21,37-46.
- Baron-Cohen, S. (1989). « The autistic child's theory of mind : a case of specific developmental delay. *Journal of Child Psychology and Psychiatry* 30 295-297.
- Baron-Cohen, S. (1995). *Mindblindness. An essay on autism and theory of mind*. Trad.fr. *La cécité mentale*. Presses Universitaires de Grenoble, 1998.
- Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention, *Consciousness and Cognition*, 17, 557-64.
- Beckmann, M., Johansen-Berg, H., & Rushworth, M.F.S (2009). Connectivity-based parcellisation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, 29, 1175-1190.
- Bedell, J., Lennox, S.S., Smith, A.D., & Rabinowicz, E.F. (1998). Evaluation of problem solving and communication skills of persons with schizophrenia, *Psychiatry Res.*, 78, 197-206.
- Blakemore, S. J., & Decéty, J. (2001). From the perception of action to the understanding of intention. *Nature Neuroscience*, 2, 561-567.
- Blakemore, S.J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The detection of contingency and animacy from simple animations in the human brain, *Cerebral cortex*, 13, 837-844.
- Blakemore S. J., & Frith, U. (2004). How does the brain deal with the social world? *Neuroreport*, 15, 119-28.
- Blakemore, S-J. *et al.* (2004). Social cognitive neuroscience: where are we heading? *Trends Cogn. Sci.* 8, 216-221.
- Bonda E, Petrides M, Ostry D, Evans A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci* 16: 3737-44.

- Brankovic, S.B., & Paunovic, V.R. (1999). Reasoning under uncertainty in deluded schizophrenic patients: A longitudinal study, *European Psychiatry*, 14, 76-83.
- Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes vs. action simulation. *Current Biology*, 17, 2117-2121.
- Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.
- Brunet, E. Sarfati, Y., & Hardy-Bayle, M.C. (2003). Reasoning about physical causality and other's intentions in schizophrenia, *Cogn Neuropsychiatry*, 8, 129-139.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400-404.
- Carey, S. (1985). *Conceptual Change in Childhood*, Cambridge, MA: MIT Press.
- Carruthers, P. (1996). *Simulation and self-knowledge: a defence of the theory-theory*. In P. Carruthers & P.K. Smith, Eds. *Theories of theories of mind*. Cambridge: Cambridge University Press.
- Castelli, F., Happé, F., Frith, U. & Frith, C. D. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314-325.
- Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., & Walter, H. (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*, 45, 3105-13.
- Colbert, S.M., & Peters, E.R. (2002). Need for closure and jumping-to-conclusions in delusion-prone individuals, *Journal of Nervous and Mental Disease*, 190, 27-31.
- Corcoran, R. *et al.* (1995). Schizophrenia, symptomatology and social inference: Investigating 'theory of mind' in people with schizophrenia. *Schizophrenia Research*, 17, 5-13.
- Corcoran, R. (2000). Theory of mind in other clinical populations. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience* (2nd ed.), pp. 391-421. Oxford: Oxford University Press.
- Corcoran, R. (2001). Theory of Mind in Schizophrenia, In Penn, D., Corrigan, P. (eds), *Social Cognition in Schizophrenia*, APA, Washington.
- Corcoran, R. (2001). Theory of mind in schizophrenia. In D. Penn & P. Corrigan (eds.), *Social cognition in schizophrenia* (pp. 149-174). Washington, DC: American Psychiatric Association.
- Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, 26, 237-67.
- Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychol*, 124, 60-78.

- Cutting, J. & Murphy, D. (1990). Impaired ability of schizophrenic, relative to manics or depressives, to appreciate social knowledge about their culture, *Br. J. Psychiatry*, 157, 355-358.
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering H (2008). Complementary systems for understanding action intentions. *Current Biology*, 25, 454-7.
- Domenech, P. & Dreher J.-C. (en revision). Decision threshold in the human brain
- Doody, G.A. *et al.* (1998). Theory of mind and psychoses. *Psychological Medicine*, 28, 397-405.
- Drury, V.M. *et al.* (1998). 'Theory of mind' skills during an acute episode of psychosis and following recovery. *Psychological Medicine*, 28, 1101-1112.
- Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Ann N Y Acad Sci* 1169, 448-58.
- Faillenot I, Toni I, Decety J, Grégoire MC, Jeannerod M. (1997). Visual pathways for object-oriented action and object recognition: functional anatomy with PET. *Cereb Cortex* 7: 77-85.
- Fazio P, Cantagallo A, Craighero L, D'Ausilio A, Roy AC, Pozzo T, Calzolari F, Granieri E, Fadiga L (2009). Encoding of human action in Broca's area. *Brain* 132, 1980-8.
- Filimon, F., Nelson, J.D., Hagler, D.J., Sereno, M.I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage* 37, 1315-1328.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J Pers Soc Psychol*, 82, 878-902.
- Fletcher, P.C. & Frith, C.D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10: 48-58.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. *Science* 308, 662-667.
- Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d'Amato, T., & Jeannerod; M. (2001). Defective recognition of one's own actions in patients with schizophrenia, *American Journal of Psychiatry*, 158, 454-459.
- Friston, K., Kilner, J., & Harrison, L. (2006). *A free energy principle for the brain*. *J Physiol Paris*, 100:70-87.
- Frith U. (1989). *Autism : explaining the enigma*. Trad. fr. *L'énigme de l'autisme*, O. Jacob, Paris, 1992.
- Frith, C.D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates.
- Frith, C., & Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychological Medicine*, 26, 521-530.

- Frith, U. et Happé, F. (1999), Theory of Mind and Self-Consciousness: What Is It Like To Be Autistic? *Mind and Language*, 14 (1), pp. 1-22.
- Frith, U. & Frith C.D. (2003). Development and neurophysiology of mentalizing. In *The Neuroscience of Social Interaction* (Frith, C. and Wolpert, D. eds.), pp. 45-75, Oxford University Press.
- Frith, C.D. (2004). Schizophrenia and theory of mind, *Psychological Medicine*, 34, 385-389.
- Frith, C.D. (2005). The self in action: Lessons from delusions of control, *Conscious Cognitive*, 14, 752-770.
- Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, 1079, 36-46.
- Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, 17, 724-32.
- Fu, G., & Lee, K. (2007). Social grooming in the kindergarten: the emergence of flattery behavior. *Developmental Science*, 10, 255-265.
- Gallagher, H.L., Frith, C.D. (2003). Functional imaging of 'theory of mind'. *Trends Cogn Sci* 7, 77-83.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 596-609.
- Gallese, V., & Goldmann, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci*, 2, 493-501.
- Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philos. Trans. R.Soc. Lond. B Biol. Sci.* 258, 517-528.
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci*, 8, 396-403.
- Gazzola, V., Rizzolatti, G., Wicker, B., Keysers, C. (2007b). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35, 1674-1684.
- Georgieff, N. (2000) Neuropsychopathologie cognitive sociale de l'action : apport à l'étude des symptômes positifs de la schizophrénie. *Intellectica* 2, 31, 191-225.
- Georgiou, I., Becchio, C., Glover, S., & Castiello, U. (2007). Different action patterns for cooperative and competitive behaviour. *Cognition*, 102, 415-433.
- Gergely, G., et al. (1995). Taking the intentional stance at 12 months of age. *Cognition*, Vol. 56, No. 2., 165-193.
- German, T.P., Niehaus, J.L., Meghan, P., Roarty, M.P., Giesbrecht, B., Miller, M.B., 2004. Neural correlates of detecting pretense: automatic engagement of the intentional stance under covert conditions. *J. Neurosci.* 16, 1805-1817.
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36, 299-308.
- Goldman, A.I. (1995). In defense of the Simulation Theory. Dans *Mental simulation: Philosophical and Psychological Essays*, édité par M. Davies et R. Stone, Oxford : Blackwell.

- Goldman, A. I., & Sebanz, N. (2005). Simulation, mirroring, and a different argument from error. *Trends Cogn Sci*, 9, 320.
- Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, 7, 145-71.
- Gopnik, A. (1993). How we know our minds : the illusion of first-person knowledge of intentionality. *The Behavioral and Brain sciences*, 16, pp. 1-14.
- Gopnik, A., & Wellman, H. (1994). The theory theory. In Hirschfeld, L. and Gelman, S. (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge: Cambridge University Press.
- Gopnik, A. & Meltzoff, A.N. (1994). Minds, bodies and persons: Young children's understanding of the self and others as reflected in imitation and 'theory of mind' research. In *Self-awareness in animals and humans*, édité par S. Parker et R. Mitchell, New York: Cambridge University Press.
- Gordon, R.M. (1996). Radical simulationism. Dans *Theories of Theories of Mind*, édité par P. Carruthers et P. Smith, Cambridge : Cambridge University Press, pp. 11-21.
- Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representation in humans by positron emission tomography: 2. Observation compared with imagination. *Experimental Brain Research*, 112, 103-111.
- Grafton, S. T., & Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum Mov Sci*, 26, 590-616.
- Grèzes, J., & Decéty, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human Brain Mapping*, 12, 1-19.
- Grèzes, J., Armony, J.L., Rowe, J., & Passingham, R.E.(2003). Activations related to 'mirror' and 'canonical' neurones in the human brain: an fMRI study. *NeuroImage* 18, 928–937.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition, In Ron Sun (ed.), *The Cambridge handbook of computational cognitive modelling*, Cambridge: Cambridge University Press.
- Happé, F.G.E. (1995). The role of verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843-855.
- Hamilton, A.C. & Grafton, S. (2006). Goal representation in human anterior intraparietal sulcus. *J Neurosci* 26, 1133–1137.
- Hamilton, A.C. & Grafton, S. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex*, 18, 1160-1168.
- Harrington, L., Siegert, R. J. & McClure, J. (2005) Theory of mind in schizophrenia: a critical review. *Cognitive Neuropsychiatry*, 10, 249 -286.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behaviour. *American Journal of Psychology*, 57, 243-259
- Herold, R. et al. (2002). Theory of mind deficit in people with schizophrenia during remission. *Psychological Medicine*, 32, 1125–1129.

- Hugues, C. *et al.* (1984). Evidence for central executive dysfunction in autism. *Neuropsychologia*, 32, 477-492.
- Huq, S.F., Garety, P.A., & Hemsley, D.R. (1988). Probabilistic judgements in deluded and non-deluded subjects, *Quarterly Journal of Experimental Psychology: Human Learning and Memory*, 40, 801-812.
- Hurlburt, R.T. *et al.* (1994), Sampling the form of inner experience in three adults with Asperger syndrome. *Psychol Med*, 24(2), pp. 385-95.
- Iacoboni, M. *et al.* (1999). Cortical mechanism of human imitation. *Science* 286, 2526-2528.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3, 529-535.
- Jabbi, M., Bastiaansen, J., & Keysers, C. (2008). A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. *PLoS One*, 13, e2939.
- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language*, 23, 190-223.
- Jacob, P. (2009). The tuning-fork model of human social cognition: a critique. *Consciousness and Cognition*, 18, 229-43.
- Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends Cogn Sci*, 9, 21-25.
- Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, 33.
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, 14, 103-109.
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation, and self-identification. *Mind and Language*, 19, 113-146.
- Jones, E. (1999). The phenomenology of abnormal belief. *Philosophy, Psychiatry and Psychology*, 6, 1-16.
- Kaliuzhna, M., Chambon, V., & van der Henst, J.-B. (submitted). Investigating hypothesis testing and advice taking in deluded patients: should we revise our beliefs about belief revision?
- Koechlin, E., Danek, A., Burnod, Y., and Grafman, J. (2002). Medial prefrontal and subcortical mechanisms underlying the acquisition of motor and cognitive action sequences in humans. *Neuron* 35, 371-381
- Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12, 939-947.
- Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends Cogn Sci*, 11, 194-6.
- Keysers, C. & Perrett, D.I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci* 8, 501-507.
- Kilner JM, Vargas C, Duval S, Blakemore SJ, Sirigu A. (2004). Motor activation prior to observation of a predicted movement. *Nat Neurosci* 7: 1299-301.

- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). Predictive coding: an account of the mirror neuron system. *Cogn Process*, 8, 159-66.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, 18, 619-23.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308, 78-83.
- Knill, D.C. & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation, *Trends Neurosci* 27: 712-9.
- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philos Trans R Soc B*, 363, 2021-31.
- Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn Sci*, 10, 319-26.
- Langdon, R., Coltheart, M., Ward, P.B., & Catts, S.V. (2002). Disturbed communication in schizophrenia: The role of poor pragmatics and poor mind-reading, *Psychological Medicine*, 32, 1273-1284.
- Lee, T.W., Josephs, O., Dolan, R.J., & Critchley, H.D. (2006). Imitating expressions: emotion-specific neural substrates in facial mimicry. *Soc Cogn Affect Neurosci*, 1, 122-135.
- Leekam, S.R., & Prior, M. (1994). Can autistic children distinguish lies from jokes? A second look at second-order belief attribution. *Journal of Child Psychology and Psychiatry*, 35, 901-915.
- Leslie, A. M. (1987). Pretense and representation: the origins of "theory of mind". *Psychologica I Review*, 94, 412-42.
- Leslie, A.M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, 1, 247–254.
- Luppino G, Murata A, Govoni P, Matelli M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Exp Brain Res* 128:181-7.
- Maguire E.A. (2001) The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand J Psychol* 42:225–238.
- Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, 81, B1-9.
- McCabe, R. (2004). On the Inadequacies of Theory of Mind Explanations of Schizophrenia. *Alternative Accounts of Alternative Problems, Theory & Psychology*, 14, 738-752.
- Manto, M. (2009). Mechanisms of human cerebellar dysmetria: experimental evidence and current conceptual bases. *J Neuroeng Rehabil*, 13, 6-10. Review.
- Mazza, M., De Risio, A., Surian, L., Roncone, R., & Casacchia, M. (2001). Selective impairment of theory of mind in people with schizophrenia, *Schizophrenia Research*, 47, 299–308.

- Mele, A. (1992). *Springs of Action: Understanding Intentional Behavior*, Oxford University Press.
- Metzinger, T. & Gallese, V. (2003). The emergence of a shared action ontology: building blocks for a theory. *Self and Action*. Special issue of *Consciousness & Cognition* (12:4), 549-571.
- Mitchell, J. P. (2005). The false dichotomy between simulation and theory-theory: the argument's error. *Trends Cogn Sci*, 9, 363-364.
- Morecraft RJ, Cipolloni PB, Stilwell-Morecraft KS, Gedney MT, Pandya DN. (2004). Cytoarchitecture and cortical connections of the posterior cingulate and adjacent somatosensory fields in the rhesus monkey. *J Comp Neurol* 26: 37-69.
- Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research*, 21, 215-223.
- Nichols, S., & Stich, S. (2003). *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds*, Oxford University Press, New York.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Onishi, K.H. & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308: 255-8.
- Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychol*, 124, 106-28.
- Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, 22, 129-151.
- Pacherie, E. (2000). The content of intentions. *Mind and Language*, 15, 400-432.
- Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition*, 107, 179-217.
- Pelphrey KA, Mitchell TV, McKeown MJ, Goldstein J, Allison T, McCarthy G. (2003). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *J Neurosci* 23: 6819-25.
- Perner, J. et al. (1989). Exploration of the autistic child's « theory of mind »: knowledge, belief and communication. *Child Development*, 60, 689-700.
- Phillips, W. (1993). Understanding intention and desire by children with autism. Thèse de doctorat non publiée, Institute of Psychiatry, University of London.
- Phillips, M.L. et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* 289, 495-498
- Picard, N. & Strick, P.L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cereb Cortex* 6:342-353.
- Picard, N. & Strick, P.L. (2001). Imaging the premotor areas. *Curr Opin Neurobiol* 11:663-672.
- Pickup, G.J. & Frith, C.D. (2001). Theory of mind impairments in schizophrenia: symptomatology, severity and specificity. *Psychological Medicine* 31, 207-220.
- Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1, 515-526.

- Puce, A. & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond B Biol Sci* 358: 435-45.
- Raffman, D. (1999), What Autism May Tell Us About Self-Awareness : A Commentary on Frith and Happé. *Mind and Language*, 14 (1), pp. 23-31.
- Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation* , University of Michigan Press.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.
- Rizzolatti G., *et al.* (2000). Cortical mechanism subserving object grasping and action recognition: a new view on the cortical motor functions. In *The New Cognitive Neurosciences* (Gazzaniga, ed.), pp. 539-552.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Neurosci Rev*, 2, 661-670.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu Rev Neurosci*, 27, 169-92.
- Rushworth, M.F. & Taylor, P.C. (2006). TMS in the parietal cortex: updating representations for attention and action. *Neuropsychologia* 44, 2700-16.
- Sakata H, Taira M, Murata A, Mine S. (1999). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb Cortex* 5: 429-38.
- Sarfati, Y., Hardy-Bayle, M.C., Nadel, J., Chevalier, J.-FR., & Widlocher, D. (1997). Attribution of mental states to others by schizophrenic patients, *Cognitive Neuropsychiatry*, 2, 1-17.
- Sarfati, Y., Hardy-Bayle, M.C., Brunet, E., & Widlocher, D. (1999). Investigating theory of mind in schizophrenia: influence of verbalization in disorganized and non-disorganized patients, *Schizophrenia Research*, 25, 183-190.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn Sci*, 10, 70-6.
- Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: the role of expertise. *Psycho Bull Rev*, 16, 70-5.
- Saxe, R. *et al.* (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *American Review of Psychology* 55, 87-124.
- Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N. (2004b). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446.
- Saxe, R. (2005). Against simulation: the argument from error. *Trends Cogn Sci*, 9, 174-9.
- Schneider, K. (1959). *Clinical Psychopathology*. New York: Grune & Stratton.
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends Cogn Sci*, 4, 299-309.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press.

- Searle, J. (1995). *The Construction of Social Reality*, New York: The Free Press.
- Seltzer B & Pandya DN (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol*. 15: 445-63.
- Singer, T. *et al.* (2004). Empathy for pain involves the affective but not the sensory components of pain. *Science* 303, 1157-1162
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466-9.
- Spiers HJ, Maguire EA (2007) The neuroscience of remote spatial memory: a tale of two cities. *Neuroscience* 149:7–27.
- Sprong, M., Schothorst, P., Vost, E., Hox, J., & van Engeland, H. (2007). Theory of mind in schizophrenia: a meta-analysis, *British Journal of Psychiatry*, 191, 5-13.
- Summerfield, C. & Koechlin E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, 59: 336-347.
- Summerfield, C. & Egner, T. (2009). Expectation (and attention) in visual cognition, *Trends Cogn Sci*, 13: 403-9.
- Surian, L., Caldi, S. & Sperber, D. (2007). Attribution of Beliefs by 13-Month-Old Infants. *Psychological Science*, 18: 580-6.
- Taylor, G.J. (1994). The alexithymia construct: conceptualization, validation, and relationship with basic dimensions of personality, *New Trends Exp Clin Psychiatry*, 10, 61–74.
- Tenenbaum, J.B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In Gopnik, A., & Schulz, L. (eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford University Press.
- Umiltà MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G. (2001). I know what you are doing. a neurophysiological study. *Neuron* 31: 155-65.
- Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Soc Neurosci* 4, 165–184.
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp* 30, 829–858.
- Van Overwalle, F. & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, in press.
- Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., Bara, B.G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J. Cogn. Neurosci.* 16, 1854–1863.
- Walter, H., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R.B., Erk, S., & Bara, B.G. (2009). Dysfunction of the social brain in schizophrenia is modulated by intention type: An fMRI type, *SCAN*, 4, 166-176.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293-1313.

- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representations and constraining functions of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.
- Wolpert, D.M. *et al.* (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc.* 358: 593-602.
- Ybarra, O. (2002). Naive causal understanding of valenced behaviors and its implications for social information processing. *Psychol Bull*, 128, 421-41.
- Zalla, T., Verlut, I., Franck, N., Puzenat, D., & Sirigu, A. (2004). Perception of dynamic action in patients with schizophrenia, *Psychiatry Research*, 128,39-51.
- Zalla, T., Bouchilloux, N., Labruyere, N., Georgieff, N., Bougerol, T., & Franck, N. (2006). Impairment in event sequencing in disorganised and non-disorganised patients with schizophrenia. *Brain Research Bulletin*, 68,195-202.

Publications

Articles soumis ou en préparation

- Chambon V.**, Barbalat G., Jacquet P., Pacherie E., Franck N., & Farrer C. (in prep.). Mindreading under influence: abnormal dependence on prior knowledge in patients with schizophrenia
- Chambon V.**, Domenech P.,Pacherie E.,Koechlin E., Baraduc P., & Farrer C. (submitted). What are they up to? The role of sensory evidence and prior knowledge in action understanding.
- Jacquet P., **Chambon V.**, Demily C., & Franck N. (submitted). "If I didn't say that, then I thought it aloud": Internal source-monitoring ambiguity and its relationship to formal thought disorder in schizophrenia.
- Barbalat G., **Chambon V.**, Ody C., Domenech P., Koechlin E, Franck N., & Farrer C. (submitted). Top-down control within the lateral prefrontal cortex in schizophrenia.
- Kaliuzhna M., **Chambon V.**, Franck, N., & Van der Henst J.-B. (submitted) Investigating hypothesis testing and advice taking in delusional patients: should we revise our beliefs about belief revision?

Articles acceptés pour publication

- Chambon V.** (à paraître). Un modèle hiérarchique des représentations d'actions intentionnelles dans la schizophrénie, *L'évolution Psychiatrique*
- Barbalat G., **Chambon V.**, Franck N., Koechlin E., Farrer C. (2009). Organization of cognitive control within lateral prefrontal cortex in schizophrenia, *Archives of General Psychiatry*, 66, 1-10.

- Demily C., Cavézian C., Desmurget M., Berquand-Merle M., **Chambon V.**, Franck N. (2009). The game of chess enhances cognitive abilities in schizophrenia, *Schizophrenia Research*, 107, 112-113.
- Chambon V.**, Franck N., Koechlin E., Ciuperca G., Fakra E., Azorin J.-M., & Farrer C. (2008). The architecture of cognitive control in schizophrenia, *Brain*, 131, 962-970.
- Chambon V.**, Baudouin J.-Y., & Franck N. (2006). The role of configural information in facial emotion recognition in schizophrenia, *Neuropsychologia*, 44, 2437-44.
- Baudouin J.-Y., **Chambon V.**, & Franck N. (2009). Expert en visage ? Pourquoi sommes-nous tous... des experts en reconnaissance des visages [Expert in face? Why are we all... expert in face recognition], *L'évolution psychiatrique*, 74, 3-25.
- Chambon V.**, Baudouin J.-Y. (2009). Reconnaissance de l'émotion faciale et schizophrénie [Facial emotion recognition in schizophrenia], *L'évolution psychiatrique*, 74, 123-135.

Chapitres d'ouvrage

- Sescousse G., Bault N., & **Chambon V.** (à paraître). « Argent, sexe et chocolat : comment le cerveau représente-t-il la récompense ? » In J.-M. Roy (ed.), *Peut-on se passer des représentations en Sciences Cognitives ?* De Boeck, Bruxelles.
- Chambon V.** (2009). « Les mécanismes de reconnaissance intentionnelle chez le sujet sain et le patient schizophrène ». In J. Rozenberg, C. Hervé & N. Franck (Eds.), *Des neurosciences à la psychopathologie : action, langage, imaginaire*, De Boeck, Bruxelles.
- Chambon V.**, Franck N., & Rozenberg J. (2009). « Syndrome d'ipséité, hallucinations et troubles du rapport d'altérité chez les patients schizophrènes. » In J. Rozenberg, C. Hervé & N. Franck (Eds.), *Des neurosciences à la psychopathologie : action, langage, imaginaire*, De Boeck, Bruxelles.
- Chambon V.**, Vernet M., Martin F., Baudouin J.-Y., Tiberghien G., & Franck N. (2007). "Visual pattern recognition: what makes faces so special?" In M.S. Corrigan (Ed.), *Pattern Recognition in Biology* (pp. 1-19). New York: Nova Publishers.