UNIVERSITÉ DE BOURGOGNE - FRANCHE-COMTÉ

ÉCOLE DOCTORALE « SOCIÉTÉS, ESPACES, PRATIQUES, TEMPS »

Thèse en vue de l'obtention du titre de docteur en

PSYCHOLOGIE

CAPACITÉ DE LA MÉMOIRE DE TRAVAIL ET SON OPTIMISATION PAR LA COMPRESSION DE L'INFORMATION

Présentée et soutenue publiquement par

Mustapha CHEKAF

Le 13 janvier 2017

Sous la direction de Fabien MATHY, Professeur

Membres du jury :

Lucile CHANQUOY,	Professeure, Université Nice Sophia Antipolis,	
André DIDIERJEAN,	Professeur, Université de Bourgogne - Franche Comté,	
Fernand GOBET,	Professor, University of Liverpool,	
Benoît LEMAIRE,	Maître de Conférences, Université Grenoble Alpes,	Rapporteur
Fabien MATHY,	Professeur, Université Nice Sophia Antipolis,	
Arnaud REY,	Chargé de recherche, CNRS & Aix-Marseille Université,	Rapporteur

Remerciements

En premier lieu j'adresse mes remerciements à Lucile Chanquoy, André Didierjean, Fernand Gobet, Benoît Lemaire et Arnaud Rey, pour avoir accepté de faire partie de mon jury de thèse, pour leur lecture, leurs commentaires et leurs conseils avisés.

Je remercie Fabien Mathy, dont l'incessante disponibilité m'impressionne encore, pour son travail, son énergie et son endurance contagieuses. Cette thèse doit tout (ou partie) aux marqueurs rouge, noir, bleu, vert, aux surligneurs jaune, orange, vert, aux agrafes, trombones et au recyclage papier.

Merci infiniment à Nicolas Gauvrit et Alessandro Guida, pour les nombreux échanges instructifs, les épuisantes corrections, les regains d'énergie sortie de nulle part. Je remercie (encore) Fabien Mathy et Nelson Cowan pour la richesse des remarques et des conseils théoriques et méthodologiques qui ont alimenté de nombreux allers-retours (d'emails).

C'est au laboratoire de psychologie de Besançon que j'ai débuté cette thèse, et je remercie tou-te-s les enseignant-e-s chercheur-e-s, particulièrement André Didierjean, redécisiviste dans mon parcours, et l'opérateur logique flou, Roland (Łukasiewicz|Gödel) Schneider qui m'a plus d'une fois donné mal au crâne. Merci à Géraldine, Sylvie et Annick pour votre bonne humeur à l'épreuve de tout. Dans ce registre, au labo et ailleurs à Besançon, mes plus drôles et touchants moments je les ai passés avec vous, Agathe, Abdelnour, Jonathan, Anaïs, Renzo, Gao, Nicolas, Lucie, Stella, Cyril, Sarah, et Delphine. Merci Johanna pour tes encouragements et ton amitié, et enfin, comme à Kigali, je colle mes tempes contre les tiennes Claude.

J'ai terminé cette thèse à Nice, au sein du laboratoire Bases, Corpus, Langage, et je veux remercier tou-te-s les membres pour leur accueil amical. Je voudrais exprimer ma reconnaissance à Elisabetta Carpitelli, Tobias Scheer, Damon Mayaffre, Lucile Chanquoy, Michèle Olivieri, pour avoir permis et facilité mon intégration parmi les doctorant-e-s, l'équipe *Langage et Cognition* et l'ensemble des membres du labo. Merci à Pierre-Aurélien, Carlos, Trinidad, Diane,

François, Katerina, Frédéric, Mickaël, Marie-Claire, Émilie, Marie-Albane, Jean-Pierre, Christian. Je veux bien entendu remercier Marwa, Samaneh, Philippe, Julia, Heba, Rosa, Sophie, Aisha, Miriam, Camille, Claire, Cécile, l'hilarante équipe as de pique, Morgane, Bruni l'épique équipe cc, Sabrina, Séverine les *Reservoir Docs*. Merci simplement à Laurent, Richard, Caro, et évidemment, Magali.

J'adresse mes remerciements à la MSH et à toutes les personnes qui font de ce grand bâtiment, dédié à la recherche et l'interdisciplinarité, un espace de richesse et de rencontre dans lequel j'ai eu la joie de partager des moments parfois improbables, en partie sur la passerelle, administrateur central de la MSH, mais aussi au Lapcos, mon autre labo d'adoption, une famille, des ami-e-s, Hanane Ramzaoui, petitAlex, Romain, Anaïs, Camille, Candice, Lou, Fernanda, Victor, Zaineb, grandAlex et Tania, l'algorithme convivial. Et comme il faut bien que l'essentiel soit fait, que soient remercié-e-s celles et ceux qui aiment alimenter débats et soirées, Dorian, Élise, Ismaël, l7aj Jamal, Loubna, Savéria, Sihem, Tom. Parmi les adeptes de la passerelle, quelques sociologues, tou-te-s engagé-e-s et passionné-e-s comme Aurélie, Giulia, Jean-Philippe, Yumiko et Arihanita. La bonne humeur doit également aux archéologues armé-e-s de café et chocolat, Alain, Arnaud, Benjamin, Christina, Gaëlle, Laura, Léa et Louise. Hors les murs, puisque la vie les dépasse, pour votre soutien, merci Pablo, John, Angus et Laetitia.

Enfin, je pousse un cri afin de passer à Table des matières, et inviter tout affectueusement mes soeurs, mes frères et mes parents, Bouchta, Fatima qui avec force, coeur, huile d'olive ont fait ce qu'ils avaient à faire.

Table des matières

Remerciements 1

I Introduction **7**

Mémoire à court terme Vs. mémoire de travail 9
Tâches de mémoire à court terme Vs. tâches de mémoire de travail 11
Pouvoir prédictif de la mémoire à court terme et de la mémoire de travail . . . 12
Chunking : liens entre mémoire à courte terme, mémoire de travail et mémoire à long terme 13
La formation de chunks en mémoire de travail 17
Chunking ou Grouping ? 18
Paradoxes de la mémoire à court terme passive 20
Relation entre (*stockage* × *traitement*) et intelligence 22
Tâches d'empan de Chunks 25
Tâches basées sur le jeu Simon et complexité algorithmique 25
Tâches de catégorisation et complexité de Feldman 29
Objectifs de la thèse 32
Resumé des articles 35
Article 1 : The Formation of Chunks in Immediate Memory and its Relation to Data Compression 35
Article 2 : There is no fundamental difference between simple and complex span tasks 37
Article 3 : Chunking in Working Memory and its Relationship to Intelligence . 38

II Formation de chunks en mémoire immédiate et compression de données **41**

Introduction 45
Method 48
Results 55
Discussion 65
Interference-based models and local distinctiveness account 68
Compressibility of information 70
Conclusion 72
Appendix 74
Notes on Feldman's (2000) compressibility metric 74

III Les tâches d'empan simples et complexes ne présentent pas de différence fondamentale 79
Experiment 1 . . . 86
Results . . . 93
Experiment 2A . . . 95
Method . . . 97
Results . . . 97
Experiment 2B . . . 99
Method . . . 100
Results . . . 101
Experiment 2C . . . 101
Method . . . 102
Results . . . 104
Discussion . . . 105
Experiment 2D . . . 108
Method . . . 108
Results and Discussion . . . 109
General Discussion . . . 109

IV Mesure de la capacité de la mémoire de travail par une tâche d'empan de chunks et relation avec l'intelligence 113
Experiment 1 . . . 119
Method . . . 120
Results . . . 125
Discussion . . . 132
Experiment 2 . . . 135
Method . . . 136
Results . . . 138
Discussion . . . 146
General discussion . . . 147

V Conclusion Générale 151

Références Bibliographiques 169

Liste des Figures 183

Liste des Tableaux 189

Première partie

Introduction

Mémoire à court terme Vs. mémoire de travail

Dès la fin du XIX^{ème} siècle, avec la démocratisation de la scolarisation des enfants, les recherches se sont intéressées aux différences de performance scolaire entre les élèves. Les premiers tests d'intelligence ont été ensuite très rapidement construits (début XX^{ème}) afin d'évaluer les capacités mentales des élèves et de les orienter vers des enseignements appropriés (Danziger, 2008). Une des premières façons d'évaluer les capacités mentales des enfants a été de mesurer le nombre de chiffres ou de lettres que l'enfant pouvait répéter après les avoir entendus. On a vite constaté que les enfants de deux écoles de Londres qui avaient de meilleurs résultats scolaires avaient tendance à se rappeler plus de chiffres ou de lettres que les élèves les plus faibles (Jacobs, 1887). Depuis, des tests d'empan variés, inspirés de ceux de Jacobs, sont inclus dans des batteries de tests d'évaluation de l'intelligence tels que la Wechsler Intelligence Scale for Children (WISC) ou la Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 2003, 2008). Ces tests d'empan consistant à retenir des suites de lettres ou de chiffres sont appelés tests d'empan simples et sont en général associés au concept de mémoire à court terme. Leur robustesse est très élevée puisque les mesures d'empan sont stables depuis 1920 (Gignac, 2015). Cependant, ce concept de mémoire à court terme ne fait pas consensus. Il existe plusieurs conceptions qui recoupent tout ou partie de notre mémoire des événements immédiats et ces conceptions renvoient à diverses terminologies et à des paradigmes expérimentaux également diversifiés. Après la brève revue de littérature suivante, nous proposerons une conception nouvelle du fonctionnement de la mémoire à court terme qui implique une capacité de traitement habituellement associée à la mémoire à long terme : le processus de chunking.

Tout d'abord, alors que les termes "mémoire immédiate" (Miller, 1956) ou "mémoire primaire" (Waugh & Norman, 1965, suivant William James, 1890) ont été utilisés jusque dans les années 1970, ils ont été progressivement remplacés par les termes "mémoire à court terme",

probablement en raison de l'influence grandissante de l'informatique (Newell & Simon, 1956), puis "mémoire de travail", principalement suite aux conceptualisations d'Atkinson et Shiffrin (1968) et de Baddeley et Hitch (1974). Le terme de mémoire de travail a probablement été utilisé pour la première fois en psychologie par Miller, Galanter, et Pribram (1960) pour décrire un ensemble organisé d'information et de procédures qu'un individu doit traiter dans le but de planifier et entreprendre des actions. Le terme de mémoire de travail a par la suite servi à décrire un système à composantes multiples qui permet de retenir et traiter simultanément un nombre limité d'information temporairement (Baddeley & Hitch, 1974). Plus tard, Baddeley (1986) a intégré dans son modèle de la mémoire de travail des processus exécutifs centraux servant à manipuler l'information stockée (e.g., mettre à jour l'information, l'inhiber, alterner entre des tâches). La mémoire à court terme est aujourd'hui communément considérée comme une composante passive de stockage de l'information alors que la mémoire de travail est souvent associée à un système central exécutif (autrement appelé administrateur central) dans lequel l'attention est une ressource partagée, dédiée à la fois au stockage et au traitement (Barrouillet & Camos, 2012 ; Cowan, 2016 ; Engle, 2002).

Ainsi, selon cette distinction qui s'est développée progressivement, la mémoire à court terme serait une conception plus ancienne faisant référence à un système passif simplement dédié au stockage, alors que la mémoire de travail serait une notion plus récente reflétant un système actif permettant à la fois stockage et manipulation de l'information, (e.g., Baddeley & Hitch, 1974 ; Colom, Rebollo, Abad, & Shih, 2006 ; A. R. Conway, Cowan, Bunting, Therriault, & Minkoff, 2002 ; Kane et al., 2004 ; Engle, Tuholski, Laughlin, & Conway, 1999 ; Aben, Stapert, & Blokland, 2012 ; voir Davelaar, 2013, pour une brève explication).

Tâches de mémoire à court terme Vs. tâches de mémoire de travail

La popularité du modèle en sous-composantes (a minima, stockage et traitement) de Baddeley et Hitch est certainement liée au paradigme de double tâche permettant une évaluation séparée des différentes sous-composantes de la mémoire de travail. Il n'en demeure pas moins que le concept de mémoire à court terme conserve son propre paradigme. La mémoire à court terme et la mémoire de travail s'évaluent selon deux types de tâches spécifiques, respectivement, les tâches d'empan simples et les tâches d'empan (voir Shipstead, Redick, & Engle, 2012). Les empans de lettres ou de chiffres (Miller, 1956) font traditionnellement référence aux tâches d'empan simples. Ces tâches d'empan simples requièrent généralement de retenir des séries d'items (chiffres, mots, images) dans l'ordre. Ces tâches sont faciles à mettre en oeuvre sous forme de tests papier-crayon par exemple. Malgré son apparente obsolescence, l'empan simple se révèle une mesure extrêmement stable et fiable depuis une centaine d'années (Gignac, 2015). En revanche, dans les tâches d'empan complexes de type double tâche, les participants doivent retenir des items tout en réalisant des tâches concurrentes (comme la fameuse tâche de suppression articulatoire, e.g., Baddeley, 2000 ; Baddeley, Thomson, & Buchanan, 1975 ; Baddeley, Lewis, & Vallar, 1984) ou en exécutant de façon discontinue des tâches concurrentes élémentaires intercalées (e.g., Barrouillet, Bernardin, & Camos, 2004 ; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007). Si le principe de la tâche complexe est plus sophistiqué, la simplicité de la tâche simple masque des qualités que nous défendrons dans cette thèse.

L'empan de lecture (Daneman & Carpenter, 1980) et l'empan d'opérations (Turner & Engle, 1989) sont deux exemples courants du paradigme de double tâche. Les tâches complexes ont présenté un intérêt majeur dans l'avancée des théories sur la mémoire de travail, notamment sur le rôle de l'attention (A. R. Conway et al., 2005 ; Ellis, 2002 ; Oberauer, 2002). En effet, les

tâches complexes peuvent permettre de déterminer comment le traitement et le stockage interviennent simultanément en sollicitant l'attention du participant (e.g., Barrouillet & Camos, 2012 ; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012a).

Quantitativement, ces tâches conduisent à des estimations des capacités en mémoire très différentes : alors que la capacité est centrée autour de 7 pour des tâches d'empan simples (Miller, 1956), elle est centrée autour de 4 pour des tâches d'empan complexes. Nous reviendrons plus loin sur cette différence d'estimation qui est ici une question centrale.

Pouvoir prédictif de la mémoire à court terme et de la mémoire de travail

Un consensus général tend à considérer que la mémoire de travail joue un rôle plus important que la mémoire à court terme dans les activités complexes (e.g., Ehrlich & Delafoy, 1990 ; Klapp, Marshburn, & Lester, 1983). Par exemple, les empans complexes prédisent mieux l'intelligence fluide que les empans simples (Cantor, Engle, & Hamilton, 1991 ; A. R. Conway et al., 2002 ; Daneman & Carpenter, 1980 ; Daneman & Merikle, 1996 ; Dixon, LeFevre, & Twilley, 1988 ; Unsworth & Engle, 2007b, 2007a ; voir aussi Colom, Rebollo, et al., 2006). C'est le cas particulièrement pour les matrices de Raven (A. R. Conway et al., 2005). Cependant, Unsworth et Engle (2007b) ont montré récemment que la prédiction de l'intelligence fluide par les tâches d'empan simples peut être améliorée en augmentant la longueur des listes. Les auteurs ont observé que lorsque la liste atteint une longueur de cinq items, les empans simples prédisent aussi bien l'intelligence que les empans complexes (voir aussi Bailey, Dunlosky, & Kane, 2011). Selon Unsworth et Engle (2007b), lorsque l'ensemble à mémoriser ou à traiter est constitué d'au moins 3 ou 4 éléments, la mémoire primaire (par similarité au concept proposé par James, 1890) n'est pas suffisante, si bien que les éléments sont déplacés en mémoire secondaire (similaire à la mémoire à long terme épisodique). Le processus de recherche et de récupération depuis

la mémoire secondaire expliquerait les points communs entre les empans simples de listes de grandes longueurs, les empans complexes et l'intelligence fluide.
Ce concept historique de mémoire à court terme ainsi que les tâches qui lui sont associées méritent donc qu'on y porte plus d'intérêt. Le présent travail inscrit la mémoire à court terme dans une conception où les processus en jeu sont bien plus actifs que dans son acceptation actuelle.

Chunking : liens entre mémoire à courte terme, mémoire de travail et mémoire à long terme

Ce que nous développons dans nos travaux, c'est que le lien entre mémoire à court terme et mémoire de travail se noue au niveau d'un processus de réorganisation de l'information appelé chunking. Ce processus faisant habilement référence à l'organisation de la mémoire à long terme explique que les listes de grandes longueurs peuvent être réorganisées par les individus afin d'être stockées (Chase & Simon, 1973 ; De Groot & Gobet, 1996 ; Miller, 1956), parce qu'elles dépassent la capacité de la mémoire immédiate. Le chunking permet d'optimiser la mémorisation en tirant avantage des régularités (stockées en mémoire à long terme) afin de réduire la quantité d'information à retenir et lui donner une forme plus maniable. Ce processus agit par segmentation ou réorganisation pour réduire le nombre d'éléments à retenir sans modifier la quantité totale d'information. Par exemple, la série de chiffres 2-3-2-4-2-5, composée de six éléments, sera plus facilement mémorisée sous la forme de trois éléments (23-24-25) stockés en mémoire à long terme.

Le chunking peut s'appuyer sur tous types de connaissances antérieures stockées en mémoire à long terme. Par exemple, la séquence 1-9-8-4 est composée de quatre chiffres pour lesquels trouver une association est complexe, et pour lesquels le recodage en deux chunks tels que 19

et 84 n'est pas optimal. Par contre pour un participant, 1984 peut éventuellement se rapporter à une date connue ou encore au titre du roman d'Orwell. Ericsson, Chase, et Faloon (1980) relatent le cas du sujet SF, capable de rappeler jusqu'à 84 chiffres. SF recodait les séquences présentées selon un système basé sur des temps de courses, ce qui lui permettait de regrouper de longues séquences en chunks. Le chunking a également été observé dans des domaines d'expertise concernant le bridge (Charness, 1979), les programmes informatiques (McKeithen, Reitman, Rueter, & Hirtle, 1981) ou encore l'électronique (Egan & Schwartz, 1979). Gobet et Simon (1996) ont étudié en détail les capacités de rappel des joueurs d'échecs confirmés. Après observation d'un échiquier pendant 5 secondes, les experts étaient capables de rappeler une cinquantaine de pièces. Cette capacité peut atteindre 160 pièces s'ils sont entraînés à ce type de tâches. Gobet (1998) estime que ces capacités dépassent très largement celles de la MCT, il émet l'hypothèse de l'existence de structures de connaissances en mémoire à long terme que les experts développeraient. Face à une configuration de jeu, cette connaissance serait activée pour permettre un encodage très rapide d'éléments. Gobet (1998) a montré que la précision du rappel de la position des pièces dépend du nombre et de la taille des chunks. La taille des chunks dépend elle-même du niveau d'expertise, les experts utilisant des chunks de tailles plus importantes. Gobet et al. (2001) distinguent finalement deux formes de chunking, la première sous contrôle stratégique, désigne un chunking délibéré, orienté vers un but, alors que la seconde serait une forme de chunking automatique et continu, qui agirait au cours de la perception. Les experts seraient capables de telles performances grâce à l'optimisation de la combinaison de ces deux formes de chunking utilisant des structures préformées en mémoire à long terme.

Toujours en lien avec la mémoire à long terme, Brady, Konkle, et Alvarez (2009) ont montré que les individus s'appuyaient également sur les régularités statistiques des stimuli pour améliorer leur mémoire de travail. Dans une tâche de mémorisation, ils ont introduit des régularités dans certaines séquences de stimuli présentés visuellement afin d'étudier l'effet de la redondance sur la mémoire de travail. Les participants ont rappelé plus d'items dans les cas où l'affichage des

stimuli respectait une régularité au fil des séquences présentées que dans les cas où il était aléatoire.

De même, Bor et collègues (Bor, Cumming, Scott, & Owen, 2004 ; Bor, Duncan, Wiseman, & Owen, 2003 ; Bor & Owen, 2007 ; Bor & Seth, 2012) ont introduit des régularités systématiques dans le matériel à mémoriser dans le but d'induire un processus de chunking. Lorsque les individus s'appuyaient sur ces régularités pour chunker, leur empan augmentait et les auteurs ont observé une plus grande activation dans les zones préfrontales latérales.

Si la capacité de la mémoire à court terme semble largement s'appuyer sur des connaissances acquises, il n'en demeure pas moins que certaines structures recodables ne sont pas nécessairement associées à une connaissance antérieure. Par exemple, la séquence 121212 n'exige pas le recours à la mémoire à long terme, son recodage simple peut se faire au vol, par une structure répétant trois fois le pattern 12. Nos travaux s'intéressent donc plus spécifiquement à ce processus de chunking lorsqu'il est centré sur les régularités présentes dans l'information à mémoriser, sans que les connaissances en mémoire à long terme ne puissent intervenir nécessairement. Le nombre 12 est certainement stocké en mémoire à long terme, mais ce n'est pas le cas de la structure globale 121212.

L'idée de Mathy et Feldman (2012) est que l'estimation de 7 ± 2 initialement trouvée par Miller (1956) correspond à une surestimation des capacités de la mémoire immédiate, causée par un processus de chunking. Si Miller (1956) parlait de 7 ± 2 chunks pour une capacité d'environ 7 ± 2 items, Mathy et Feldman (2012) envisagent plutôt une capacité de 3 ou 4 chunks (en accord avec Cowan, 2001, 2005) pour une capacité moyenne d'environ 7 ± 2 items. L'idée clé de Mathy et Feldman (2012) est que les 3 ou 4 chunks pourraient être formés au vol, c'est à dire à partir de l'information en présence, sans nécessairement recourir à la mémoire à long terme. À l'origine, la différence entre l'estimation de Cowan et celle de Miller dépend du type de tâche utilisée pour la mesurer. Selon Cowan, quatre est la capacité généralement estimée lorsque ni les

régularités, ni la mémoire à long terme ne peuvent être utilisées pour recombiner les items. En d'autres termes, la capacité est de quatre items quand on empêche les sujets de chunker. Dans le cas des expériences de Miller, le chunking de lettres ou de chiffres n'a pas été contrôlé, il est par conséquent difficile de savoir précisément ce que représente ce nombre sept. Par exemple, dans une liste de huit chiffres à mémoriser, en recodant les huit chiffres en paires, il suffit de retenir quatre chunks. Ainsi, Mathy et Feldman (2012) ont estimé la compressibilité typique d'une séquence de chiffres aléatoires à 7/4, ce qui correspond au ratio entre les estimations de Miller et de Cowan. En conclusion de cette analyse, il apparait évident que le concept de mémoire à court terme n'est pas un processus passif puisque c'est à travers lui que pourrait s'opérer un recodage actif des éléments.

Les présents travaux poursuivent cette ligne de recherche en démontrant que les participants exposés à des séquences simples ont un rappel plus élevé lorsque celles-ci présentent des régularités, indépendamment de connaissances particulières déjà présentes en mémoire à long terme (en tous cas pour les séquences elles-mêmes, en effet, les chiffres, les couleurs ou les formes élémentaires constituant les séquences sont évidemment supposées être disponibles en mémoire à long terme pour tout individu sain). Il s'agit donc dans ces travaux d'induire la formation de chunks en mémoire immédiate, tout en évitant les effets d'apprentissage à long terme des études précédentes (Bor et al., 2003, 2004 ; Bor & Owen, 2007 ; Mathy & Feldman, 2012). En effet, Mathy et Feldman (2012) par exemple, ont utilisé un protocole basé sur la mémorisation de séries de chiffres. L'analyse des chunks potentiellement exploitables dans les séquences d'items à mémoriser a pris en compte l'identification de structures élémentaires de types répétitions ou suites logiques telles que '2468'. Cependant, au delà de régularités intrinsèques, de nombreuses séquences de ce type pourraient faire partie d'éléments déjà présents dans la mémoire à long terme des participants (voir Jones & Macken, 2015). Notre travail a donc consisté à induire la formation de chunks en mémoire à court terme, sans exposer les participants à des répétitions de séquences identiques, et en limitant le risque que ces chunks aient été préalablement encodés

en mémoire à long terme.

La formation de chunks en mémoire de travail

Les recherches sur la récupération de chunks ont eu un impact considérable sur les études portant sur le rappel immédiat (Boucher, 2006 ; Cowan, Chen, & Rouder, 2004 ; Gilbert, Boucher, & Jemel, 2014 ; Gilchrist, Cowan, & Naveh-Benjamin, 2008 ; Guida, Gobet, Tardieu, & Nicolas, 2012 ; Maybery, Parmentier, & Jones, 2002 ; Ng & Maybery, 2002), mais moins de connaissances ont été produites sur le rôle de la mémoire immédiate dans la création de chunks en situations nouvelles (Feigenson & Halberda, 2008 ; Kibbe & Feigenson, 2014 ; Moher, Tuerk, & Feigenson, 2012 ; Solway et al., 2014) ou sur la réciprocité de la mémoire de travail et du chunking (Rabinovich, Varona, Tristan, & Afraimovich, 2014). La présence de régularités dans le matériel à mémoriser produit immédiatement une surestimation des capacités (e.g., Brady et al., 2009 ; Brady, Konkle, & Alvarez, 2011 ; Brady & Tenenbaum, 2013 ; Farrell, 2008, 2012). De nombreux efforts ont donc été faits pour empêcher le regroupement d'information dans les tâches de mémoire de travail pour obtenir une estimation rigoureuse de l'empan (Cowan, 2001). Seules quelques études ont manipulé le chunking, en utilisant par exemple des co-occurrences de mots apprises, des séquences structurées par une grammaire artificielle (e.g., Chen & Cowan, 2005 ; Cowan et al., 2004 ; Gilchrist et al., 2008 ; Majerus, Perez, & Oberauer, 2012 ; Naveh-Benjamin, Cowan, Kilb, & Chen, 2007), ou des chunks multi-mots (Cowan, Rouder, Blume, & Saults, 2012). Cependant, dans ce type de méthode, des séquences contenant des répétitions de patterns sont présentées lors d'une première phase. Les participants acquièrent (explicitement ou implicitement) des représentations à long terme dont ils peuvent bénéficier a posteriori dans les tâches d'empan. Cette méthode implique un processus de reconnaissance des patterns précédemment rencontrés (e.g., Botvinick & Plaut, 2006 ; Burgess & Hitch, 1999 ; Cumming, Page, & Norris, 2003 ; French, Addyman, & Mareschal, 2011 ; Robinet, Lemaire, & Gordon, 2011 ;

Szmalec, Page, & Duyck, 2012). Notre travail se différencie par une approche du processus de chunking dédié au traitement de l'information nouvelle à traiter, sans recours à un quelconque apprentissage à long terme préalable. Nous cherchons à étudier un processus de chunking qui opère au vol, sans apprentissage spécifique ou statistique. Aussi, notre utilisation du terme chunking diffère de la définition adoptée par certains chercheurs, selon laquelle le chunking doit être défini en lien avec des associations en mémoire à long terme (Cowan, 2001). Nous reprenons la conception du chunking inspirée par Mathy et Feldman (2012) et Mathy et Varré (2013) dans laquelle un processus de chunking peut opérer à court terme pour permettre des regroupements entre les éléments à mémoriser lorsque des régularités sont présentes. Cette conception reste cependant fondamentalement en accord avec la définition de Gobet et al. (2001), selon laquelle un chunk est un "ensemble d'éléments ayant de fortes associations entre eux, mais des associations plus faibles avec les éléments d'autres chunks". L'originalité de notre approche est d'étudier un processus de chunking permettant d'observer la naissance d'un chunk sans qu'une consolidation à long terme n'ait eu besoin d'opérer.

Chunking ou Grouping ?

Il n'y a pas consensus clair sur la distinction entre chunking et grouping ou encore sur le mode perceptuel ou conceptuel du chunking (Gilbert et al., 2014). Feigenson et Halberda (2008), par exemple, ont distingué une forme de chunking qui requiert un recodage conceptuel (par exemple, découper PBSBBCCNN en PBS-BBC-CNN, sur la base de concepts existants en mémoire à long terme) d'une seconde forme de chunking qui requiert seulement un processus de recodage s'appuyant sur les éléments en présence. Ils ont développé l'idée que la liste "aubergine, tournevis, carotte, artichaut, marteau, pinces" est plus facile à retenir que "aubergine, brocoli, carotte, artichaut, concombre, courgette", parce qu'elle peut être découpée en trois unités de deux types conceptuels sans qu'il soit nécessaire de référer à un concept pré-existant

d'aubergine-carotte-artichaut. Ici, le processus de recodage est basé sur la sémantique, mais les auteurs ont développé une troisième définition du chunking qui fonde le processus de découpage sur l'information spatio-temporelle (un exemple classique est de diviser les numéros de téléphone en groupes par leur proximité spatiale). Cette dernière idée est similaire à celle de Mathy et Feldman (2012) qui ont tenté de quantifier un processus de chunking immédiat en introduisant systématiquement dans le mémorandum des patterns séquentiels qui ont peu de liens avec des connaissances en mémoire à long terme. Cette conception des chunks a donc toujours un rapport étroit avec la notion originelle de ré-encodage de l'information. Dans les tâches inspirées du jeu Simon qui sont décrites plus loin, bien que les couleurs soient des concepts présents en mémoire à long terme, la séquence jaune-rouge-jaune-rouge peut facilement être codée sans connaissance conceptuelle spécifique pour former une nouvelle représentation plus compacte comme "deux fois jaune-rouge".

Des manipulations similaires à celles que nous présentons ont déjà été faites dans la littérature sur le rappel sériel, mais elle réfèrent à un processus de grouping (comme les associations induites entre des paires de mots, du grouping temporel ou spatial, voir Anderson & Matessa, 1997 ; Frankish, 1985 ; Hitch, Burgess, Towse, & Culpin, 1996 ; Henson, Norris, Page, & Baddeley, 1996 ; Ng & Maybery, 2002 ; Maybery et al., 2002 ; Ryan, 1969). La notion de 'grouping' est ambiguë car ce terme est également utilisé dans le domaine de la perception en référence à un processus de bas niveau qui permet l'organisation en patterns d'ensembles d'objets visuels (Feldman, 1999). Le grouping est supposé être automatique et pré-attentif, alors que dans la littérature sur le rappel sériel il est le plus souvent considéré comme conscient et délibéré.

Aussi, en utilisant la terminologie d'empan de chunk (plutôt que de grouping), nous souhaitons introduire une différenciation entre grouping et chunking. Le grouping est un processus dans lequel des groupes d'objets sont séparés (spatialement ou temporellement) alors que dans le chunking, des associations sont faites au sein de groupes d'objets en s'appuyant, au delà de divisions temporelles ou spatiales marquées, sur des régularités entre les objets. Dans les ex-

périences que nous proposons, l'objectif est de favoriser le chunking en manipulant plutôt les possibilités de regroupements intra-groupes que de séparations inter-groupes.

Paradoxes de la mémoire à court terme passive

Même si un consensus général assimile les tâches d'empan simples exclusivement au stockage, et les tâches d'empan complexes à une combinaison *stockage* + *traitement* de l'information, la réalité est beaucoup plus complexe et révèle des concepts entrelacés. Selon Oberauer, Lange, et Engle (2004, p. 94), il pourrait être plus pertinent de changer les choses ainsi : "empan simple = empan complexe + mécanismes ou stratégies spécialisés". La distinction entre les concepts auxquels les tâches réfèrent (mémoire à court terme et mémoire de travail) reste parfois vague, malgré quelques tentatives de clarification (Aben et al., 2012 ; Davelaar, 2013).

Une idée dominante dans de nombreuses études est que le plus grand avantage des tâches d'empan complexes est de fournir une meilleure mesure de la combinaison *stockage+traitement* que les tâches d'empan simples. La raison est que le traitement est détourné vers la tâche concurrente et que le stockage est par conséquent dénué de traitement. Cependant, l'analyse des tâches complexes s'intéresse rarement à la composante de traitement qui représente pourtant une composante essentielle du concept de mémoire de travail (Aben et al., 2012 ; A. R. Conway et al., 2005 ; Unsworth, Redick, Heitz, Broadway, & Engle, 2009).

La présentation simpliste de la mémoire à court terme conduit à un certain décalage entre sa conception et sa mesure. Le paradoxe que nous soulevons tient en partie du fait que d'une part, la mémoire à court terme est considérée comme un espace exclusivement dédié au *stockage* passif d'éléments, mais que d'autre part, lors des tâches d'empan simples qui la mesurent, les participants peuvent traiter l'information à mémoriser (sans que rien ne les empêche de faire des regroupements). Si les tâches d'empan simples n'empêchent pas le traitement, alors elles

l'incluent sans le caractériser. La mémoire de travail est quant à elle définie comme étant principalement une combinaison *stockage* + *traitement*, alors que les tâches qui la mesurent se focalisent exclusivement sur l'analyse du *stockage* en écartant l'analyse du *traitement*. Étant donné les limitations de ressources, diviser l'attention entre deux tâches concurrentes conduit souvent à une diminution des performances au moins dans une des deux tâches, sinon dans les deux. Plummer et Eskes (2015) suggèrent de mesurer un indice d'interaction reflétant l'effet réciproque des deux tâches. Cependant, dans le paradigme de double tâche, la tâche concurrente (généralement de faible niveau de difficulté) n'est que rarement prise en compte dans les méthodes de cotation (Unsworth et al., 2009). Le paradoxe, pour résumer, est que les tâches d'empan complexes censées mesurer stockage et traitement, ne mesurent objectivement qu'une capacité de stockage, alors que les tâches d'empan simples que l'on pense réduites à mesurer le stockage, mesurent le résultat de l'interaction *stockage* × *traitement*, sans que l'on puisse déterminer les parts respectives du stockage et du traitement. Ce paradoxe est à l'origine de nos interrogations sur la distinction conceptuelle entre mémoire à court terme et mémoire de travail et sur les tâches supposées incarner ces concepts. Les tâches que nous présenterons permettent un glissement, d'une conception centrée sur le stockage vers une conception contrôlée de l'interaction *stockage* × *traitement*.

D'autres tâches de mémoire de travail se concentrent aussi sur la combinaison stockage + traitement, comme par exemple la tâche d'empan envers, les séquences lettres-chiffres du subtest de la WAIS-IV (Wechsler, 2008), les tâches de type n-back, ou le running span. Même si dans ces dernières, la composante de traitement est tournée vers les éléments stockés (à la différence de la double tâche), elle ne vise pas l'optimisation du stockage. C'est cette capacité du traitement à interagir avec le stockage dans le but de l'optimiser qui nous a conduit à proposer la notion d'*empan de chunks*.

Dans les tâches d'empan envers, pour poursuivre l'exemple, la composante de traitement ne peut pas être étudiée indépendamment de la composante de stockage, en effet, le nombre d'items

à traiter est linéairement lié au nombre d'items à stocker. Par conséquent, il n'y a pas dans ce cas de séparation claire entre traitement et stockage car les deux dépendent d'un nombre similaire d'items à mémoriser. Comme nous le verrons, les tâches que nous proposons permettent d'augmenter ou de diminuer la charge de traitement indépendamment du nombre d'éléments à mémoriser en faisant varier la complexité des listes d'items, ou inversement, faire varier le nombre d'items en maintenant un même niveau de complexité. C'est uniquement dans ce cas qu'il est possible d'étudier indépendamment stockage et traitement pour évaluer leur potentialité respective et leur interaction. Par exemple, bien que la séquence "1223334444" compte 10 items à traiter séquentiellement, un seul chunk peut être stocké une fois que l'on a reconnu le modèle régulier qui rend la séquence facile à retenir (un '1', deux '2', trois '3', quatre '4'). Ce n'est pas le cas pour la séquence "8316495207", de 10 items également, dans laquelle les chunks sont plus difficiles à former. Ainsi, notre idée était de développer des tâches pour mesurer l'empan mnésique dans lesquelles stockage et traitement peuvent être mesurés simultanément et indépendamment, et nous soutenons que cela peut être fait en manipulant la complexité et en induisant un processus de chunking reflétant une relation stockage $\times$ traitement.

Relation entre ($stockage \times traitement$) et intelligence

De nombreuses études utilisant des variables latentes ont suggéré que la capacité en mémoire de travail rend compte de 30 à 50% de la variance en intelligence fluide (g) (A. R. Conway, Kane, & Engle, 2003 ; A. R. Conway et al., 2005). En comparaison, les tâches d'empan simples rendent compte d'une plus faible covariance (Shipstead et al., 2012). Encore une fois, une exception d'une importance cruciale pour nos travaux est que cet écart en termes de variance n'est plus vrai lorsque les tâches d'empan simples utilisent des séquences longues, c'est à dire de plus de 5 items (Unsworth & Engle, 2006, 2007b ; voir aussi Bailey et al., 2011 ; Unsworth & Engle, 2007a). Par conséquent des empans simples peuvent potentiellement rendre compte d'un même

pourcentage de variance de g que les empans complexes. Cette limite de 5 items est cohérente avec la limite de capacité de 4 ± 1 items mesurée par Cowan (2001). Au delà de cette limite, les participants doivent développer des stratégies d'encodage pour retenir plus d'information. Ces stratégies de compression de l'information peuvent faire intervenir des connaissances à long terme mais elles peuvent également se fonder sur la structure même de l'information à mémoriser. Ainsi, au delà de 5 items, lorsque les participants sont contraints de dépasser leur capacité brute, la complexité de l'information joue un rôle majeur si les participants utilisent une stratégie de compression uniquement basée sur l'information à mémoriser. Considérons les séquences qui dépassent cette limite de 4 ± 1 items en distinguant les séquences par leur complexité (i.e., leur compressibilité). Les séquences les plus complexes résistent à une compression importante et par conséquent, saturent la composante de *stockage*. En revanche, les séquences les moins complexes favorisent l'apparition de processus de chunking via une réorganisation de l'information effective et devraient donc impliquer une interaction de type *stockage* $\times$ *traitement*. Dans ce contexte, la variation de la complexité est importante parce qu'il est supposé que les séquences les plus complexes ne peuvent pas être facilement réorganisées. Une grande complexité correspond à des possibilités de traitement réduites. Une mesure de complexité donnée ne permettra pas d'inclure toutes les possibilités de compression.

Néanmoins, il est toujours possible que des participants utilisent des méthodes de simplification qui échappent à toute mesure de complexité. Par exemple, pour mémoriser la séquence de couleurs "rouge-bleu-jaune-vert", un participant pourrait simplifier la tâche en associant la séquence au drapeau national mauricien. Bien qu'il y ait une faible chance que les séquences à forte complexité puissent être simplifiées, les séquences les plus compressibles (i.e., de faible complexité) ont une plus grande probabilité d'être simplifiées pour optimiser leur mémorisation. Les séquences les plus simples sollicitent une grande charge de traitement parce qu'elles induisent une optimisation de la mémoire en raison des régularités qu'elles présentent. Les séquences de faible complexité sont à rapprocher d'une certaine manière aux tâches d'empan

complexes et les séquences les plus complexes, aux tâches d'empan simples et, évidemment, nous nous attendons à des empans plus faibles pour les séquences les plus complexes.

Cette différence entre du matériel simple et du matériel complexe a déjà été relevée par Bor et collègues (Bor et al., 2003, 2004 ; Bor & Owen, 2007). En comparant l'activité cérébrale au cours d'un empan simple avec du matériel chunkable (présentant des régularités systématiques) versus du matériel moins chunkable, ils ont observé une augmentation d'activité dans les zones préfrontales latérales en présence de matériel chunkable, ce que les auteurs ont interprété comme l'effet d'une réorganisation de l'information. De plus, ce pattern d'activation a également été relié à l'intelligence. De même, l'activation d'un réseau neural similaire a également été observée lors de tâches de type "n-back", empan complexes, et de raisonnement fluide (Bor & Owen, 2007 ; Colom, Jung, & Haier, 2007 ; Duncan, 2006 ; Gray, Chabris, & Braver, 2003). Selon Duncan, Schramm, Thompson, et Dumontheil (2012, p. 868), cela montre l'implication possible de l'intelligence fluide dans les processus organisateurs de détection et d'utilisation de nouveaux chunks.

À partir de cette idée, nous prédisons que l'empan mnésique pour les séquences les plus compressibles (i.e., les moins complexes) devrait mieux corréler avec l'intelligence, car c'est dans cette condition que les participants peuvent créer de nouveaux chunks afin d'optimiser leur capacité de stockage. Un autre argument dans ce sens est que les tests d'intelligence requièrent généralement traitement et stockage de l'information de façon conjointe, comme c'est le cas dans les matrices de Raven. Par conséquent, c'est lorsque les séquences sont moins complexes que l'on devrait atteindre une meilleure corrélation avec l'intelligence, puisque c'est dans cette condition que le construit $stockage \times traitement$ a un plus grand espace d'expression.

Considérer l'interaction entre *stockage* et *traitement* comme centrale dans la mesure de l'empan en mémoire de travail est à contre-courant des efforts produits dans la littérature de séparer les mécanismes de stockage et de traitement (Cowan, 2001). Cependant, nous croyons que les

hauts niveaux de cognition dépendent directement d'une capacité qui découle de l'interaction *stockage* × *traitement* plutôt que d'une dissociation entre les deux. Aussi, selon notre conception, la distinction entre mémoire à court terme et mémoire de travail n'a pas lieu d'être retenue. C'est simplement la possibilité de compresser l'information avec du matériel plus simple et du temps disponible qui augmente l'empan mnésique en mémoire immédiate.

Tâches d'empan de Chunks

Nous avons développé plusieurs tâches originales permettant de mesurer l'empan en utilisant le chunk comme une unité d'information compressée ou recodée. Ces tâches, présentées en détail plus loin, peuvent être répertoriées selon deux types. D'une part les tâches utilisant des séquences de couleurs répétées, et fondées sur une mesure de complexité algorithmique adaptée aux séquences courtes (i.e., de faible complexité). D'autre part les tâches inspirées des tâches de catégorisation utilisant des stimuli variant selon leurs formes, tailles et couleurs, fondées sur une échelle de complexité logique, la complexité de Feldman (2000).

Tâches basées sur le jeu Simon et complexité algorithmique

Cette tâche d'empan de chunks est basée sur le jeu SIMON®, un jeu familial de mémoire grand public des années 80 commercialisé sous le format d'un appareil à quatre touches de couleurs (rouge, vert, jaune, bleu). Les touches de couleurs s'allument au hasard, et le jeu consiste à reproduire la séquence de couleurs produite, en appuyant sur les boutons correspondants. Chaque fois que le joueur reproduit correctement une séquence, le nombre de couleurs à mémoriser augmente par ajout d'une couleur supplémentaire à la fin de la séquence précédente. Le jeu progresse jusqu'à ce que le joueur fasse une erreur. Gendle et Ransom (2009) ont utilisé

le Simon lors d'une procédure de mesure d'empan en mémoire de travail et ont montré que cette procédure était résistante aux effets de l'entrainement. D'autres études ont montré que ce dispositif était intéressant dans l'évaluations de différentes populations présentant des troubles de l'audition ou du langage (C. M. Conway, Karpicke, & Pisoni, 2007 ; Humes & Floyd, 2005 ; Karpicke & Pisoni, 2000, 2004). Nos études présentaient deux différences importantes avec le jeu original. Tout d'abord, les séquences n'étaient pas présentées par ajout progressif de couleurs mais entièrement en une seule présentation. Par exemple, plutôt que l'ajout progressif de couleurs comme dans le jeu original "1) bleu, 2) bleu-rouge, 3) bleu-rouge-rouge, etc.", ce qui correspond à trois séries croissantes d'une même séquence jusqu'à ce qu'une erreur ait été commise, l'adaptation proposait dès le départ la séquence "bleu-rouge-rouge". Si le rappel était correct, une nouvelle séquence était donnée. D'autre part, aucun son n'était associé aux couleurs contrairement au jeu original. Enfin, lors de la phase de réponse, quatre boutons étaient affichés à l'écran, avec l'assignation aléatoire d'une couleur parmi les quatre à chacun des boutons, afin d'éviter d'éventuelles stratégies visuo-spatiales d'apprentissage.

Les tâches que nous proposons sur ce format se basent sur l'idée que des séquences de couleurs contiennent des régularités qui peuvent donner lieu à des processus de chunking par compression de l'information. Les individus ont tendance à réorganiser l'information en chunks afin d'optimiser la quantité d'information à retenir dans l'espace disponible (Anderson, Bothell, Lebiere, & Matessa, 1998 ; Cowan et al., 2004 ; Ericsson et al., 1980 ; Logan, 2004 ; Miller, 1956, 1958 ; Naveh-Benjamin et al., 2007 ; Perlman, Pothos, Edwards, & Tzelgov, 2010 ; Perruchet & Pacteau, 1990 ; Tulving & Patkau, 1962).

On peut distinguer deux processus différents dans le chunking : la création de chunks et la récupération de chunks (Guida, Gobet, & Nicolas, 2013 ; Guida et al., 2012). Le premier intervient lorsque les individus n'ont pas une connaissance solide de l'information à traiter. Au cours de la création de chunks, les individus utilisent leur attention (Oakes, Ross-Sheehy, & Luck, 2006 ; Wheeler & Treisman, 2002) pour lier des éléments distincts, et réorganiser l'information

qu'ils traitent en chunks. Ce processus a été intégré dans différents modèles. Chen et Cowan (2009), par exemple, ont suggéré qu'une fonction cruciale du focus attentionnel (e.g., Cowan, 2001, 2005) est effectivement de permettre des associations ; le buffer épisodique de Baddeley (Baddeley, 2000, 2001) a également été avancé pour expliquer les associations. Mais dès lors que les individus ont les connaissances nécessaires pour reconnaître des groupes d'éléments (e.g., "f", "b", "i"), ils n'ont pas besoin de plus de ressources supplémentaires pour créer des chunks, mais seulement pour les récupérer en mémoire à long terme. Le processus de compression que nous envisageons d'étudier ne concerne que la création de chunks.

Une partie de notre travail a pour but de fournir une estimation précise de la compressibilité de séquences à mémoriser. Il faut noter que, contrairement aux études précédentes (e.g., Burgess & Hitch, 1999) qui ont identifié des effets de regroupements par la présence de mini courbes de position sérielle pendant le rappel (une autre possibilité est d'étudier les probabilités d'erreurs de transitions, N. F. Johnson, 1969a), nous visons ici uniquement à estimer les possibilités de chunking pour l'ensemble de la séquence à mémoriser, et de fait, nous ne focalisons pas toujours sur la façon dont les items peuvent être regroupés séquentiellement.

Les opportunités de chunking peuvent être définies comme la probabilité (ou le potentiel) d'une séquence d'être ré-encodée, compressée de sorte à pouvoir être retenue, en s'appuyant sur une réorganisation de l'information en blocs pertinents, plutôt qu'en retenant des items indépendants (la séquence 011110, par exemple, peut être recodée en 0-11-11-0 plutôt que 0-1-1-1-1-0 en utilisant une simple séparation linéaire des items, 0-2*11-0 en regroupant les deux blocs similaires, ou encore 011-110 en utilisant la symétrie, etc.). L'estimation de compressibilité que nous développons ici permet de capturer toute sorte de régularités dans les séquences qui peuvent être utilisées par les participants pour simplifier le processus de rappel.

Plus de complexité signifie simplement moins d'opportunités de chunking, ce qui indique également que la mémorisation doit être principalement basée sur la capacité de stockage. Moins de

complexité signifie qu'une séquence peut être ré-encodé pour optimiser le stockage et dans ce cas, le traitement prend la priorité sur le stockage. Nous croyons que la notion de compression est précieuse si l'on veut mettre en rapport le chunking et l'intelligence (voir Baum, 2004, ou Hutter, 2005, qui ont développé des idées similaires en intelligence artificielle), étant donnée le contraste entre une capacité à réaliser de nombreuses activités mentales complexes et une très faible capacité de stockage. La notion de complexité (algorithmique) en informatique a été développé par Kolmogorov (1965) et Chaitin (1966). Cette notion relie compression et complexité par une définition unique : la complexité algorithmique d'une séquence est la longueur du programme le plus court capable de reconstruire la séquence (Chater & Vitányi, 2003). La complexité algorithmique de longues chaînes peut être estimée et cette estimation a déjà été appliquée à différents domaines (par exemple, en génétique, Ryabko, Reznikova, Druzyaka, & Panteleeva, 2013 ; Yagil, 2009 ; en neurologie, Fernández et al., 2011, 2012 ; Machado, Miranda, Morya, Amaro Jr, & Sameshima, 2010). Contrairement aux longues chaînes, la complexité algorithmique des chaînes courtes (3-50 symboles ou valeurs) ne pouvait pas être estimée avant les développements récents de l'informatique, grâce auxquels il est maintenant possible d'obtenir une estimation fiable de la complexité algorithmique de chaînes courtes (Delahaye & Zenil, 2012 ; Soler-Toscano, Zenil, Delahaye, & Gauvrit, 2013, 2014). La méthode a déjà été utilisée en psychologie (e.g., Kempe, Gauvrit, & Forsyth, 2015 ; Dieguez, Wagner-Egger, & Gauvrit, 2015), la complexité algorithmique est corrélée à la perception humaine du hasard (Gauvrit, Soler-Toscano, & Zenil, 2014), et nous supposons qu'elle est une approximation satisfaisante des possibilités de chunking. Elle est maintenant implémentée sous forme de package sur le logiciel R, sous le nom ACSS (Algorithmic Complexity for Short Strings ; Gauvrit, Soler-Toscano, & Zenil, 2014).

L'idée à la base de la complexité algorithmique pour les chaînes courtes (ACSS) est de tirer parti du lien entre complexité algorithmique et probabilité algorithmique. La probabilité algorithmique $m(s)$ d'une chaîne s est définie comme la probabilité qu'un programme déterministe

choisi au hasard et exécuté sur une machine universelle de Turing produise s puis s'arrête. Cette probabilité est liée à la complexité algorithmique par le théorème de codage algorithmique qui stipule que $K(s) \sim -\log_2(m(s))$, où $K(s)$ est la complexité algorithmique de s.

Au lieu de choisir des programmes aléatoires sur une machine de Turing donnée, on peut également choisir des machines de Turing aléatoires et les faire tourner sur une bande vierge. Cela a été fait sur une énorme quantité de machines de Turing (plus de 10 milliards de machines de Turing), et a conduit à une distribution d de chaînes, se rapprochant de m. La complexité algorithmique pour les chaînes courtes d'une chaîne s, ACSS(s) est défini comme par $-\log_2(d)$, qui est une approximation de $K(s)$ par l'utilisation du coding theorem (voir Gauvrit, Singmann, Soler-Toscano, & Zenil, 2015). Cette méthode indique simplement qu'une chaîne s est simple lorsqu'elle peut être produite par un grand nombre de machines aléatoires.

Tâches de catégorisation et complexité de Feldman

Certaines des tâches que nous avons développées se sont basées sur une mesure de compressibilité logique inspirée de la complexité algorithmique. À partir des travaux de Feldman (2000) sur la compressibilité logique, nous avons extrait des listes d'objets plus ou moins faciles à apprendre. Par construction nous connaissons la complexité de chacune des listes d'objets que nous avons construites puisque Feldman donne une description de la règle logique qui les sous-tend et a validé globalement le lien entre complexité logique et complexité subjective (chez Feldman, la difficulté à apprendre des règles de catégorisation).

Les travaux de Feldman (2000) suggèrent que la difficulté d'une liste d'objets est directement proportionnelle à sa complexité booléenne, c'est à dire à la longueur du plus petit équivalent logique d'une proposition. Prenons l'exemple de la séquence d'objets : ■ ▲ □ △ ; petit carré noir, petit triangle noir, petit carré blanc, petit triangle blanc. Cette séquence peut être simplifiée

en retenant l'abstraction du trait commun aux quatre objets : "petit". Si ces quatre objets sont des exemples positifs d'un concept à découvrir, alors la règle "SI petit ALORS positif" permet de séparer ces objets des objets négatifs (△ ▲ □ ■ ; grand triangle blanc, grand triangle noir, grand carré blanc et grand carré noir). Pour ce concept, il n'est donc pas nécessaire d'apprendre par cœur la liste complète des objets positifs puisqu'une règle simple permet de les identifier, leur compression en la nouvelle représentation "petit" est une simplification.

Pour les expériences que nous avons développées sur ce principe, les stimuli varient selon trois dimensions binaires pour chacune. On peut représenter schématiquement ces items sur une structure en trois dimensions, un diagramme de Hasse, dont les trois dimensions (hauteur, largeur et profondeur) sont associées à celles des objets à mémoriser (taille, couleur et forme). Les objets sur chacun des huit sommets correspondent à une des huit combinaisons possibles résultant de l'opération : 2 couleurs × 2 tailles × 2 formes (Fig. 1a). Deux objets situés sur des sommets adjacents, ont une dimension de différence. Par exemple, un grand carré noir et un grand triangle noir ne diffèrent que par leur forme, ces deux objets ont la plus grande similarité possible entre eux. La distance physique la plus courte entre deux objets (en suivant les arêtes) équivaut au nombre de différences qui les séparent. Par exemple, grand triangle noir et petit carré blanc diffèrent par la forme, la taille et la couleur, et sur le diagramme, trois arêtes les séparent physiquement, ces deux objets présentent la plus grande dissimilarité possible entre eux.

Considérons les concepts simples représentés en Fig. 1b et Fig. 1c, le premier présente deux objets, le second en présente quatre. Pour décrire littéralement, sans compression un de ces ensembles, il faudrait décrire tous les objets qui le constituent selon chacune de ses trois dimensions (par exemple, pour □ ■ □ ■ △ ▲ : "grand carré blanc, grand carré noir, petit carré blanc, petit carré noir, petit triangle blanc et petit triangle noir"). La description de chacun des deux concepts de cet exemple peut être réduite à deux caractéristiques seulement, ainsi, pour décrire □ ■ □ ■ △ ▲ il suffit de retenir "carrés ou petits" et pour l'ensemble △▲ la formule

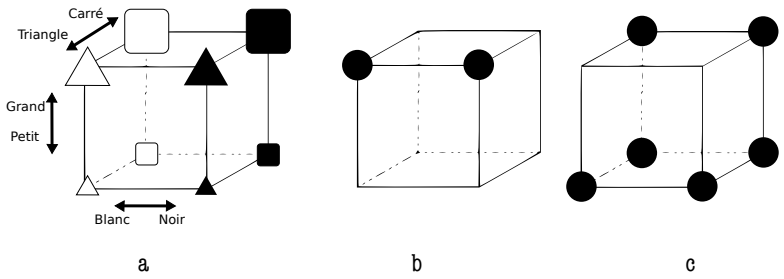


FIGURE 1 – (a) : Dans ce diagramme de Hasse, chaque sommet correspond à un objet, chaque objet est défini par trois dimensions, forme, taille, couleur. Pour l'exemple, nous avons choisi les formes (carré et triangle), les couleurs (noir et blanc) et les positions sur le diagramme (grands en haut, noirs à droite, triangles au premier plan). Cependant, pour les expériences développées plus loin, les formes, les couleurs et les positions sont tirées au sort. (b) et (c) : exemple de deux concepts, en se référant à l'exemple (a), les objets pointés en (b) correspondent à "triangles et grands" et les objets pointés en (c) correspondent à "petits ou carrés".

"triangles et grands" est suffisante. C'est la longueur de la formule la plus courte pour décrire un ensemble d'objet, ou un concept, qui est retenue comme indice de complexité. Les formules pour décrire les ensembles montrés en Fig. 1b et Fig. 1c sont de même longueur, leur indice de complexité est donc de 2.

Ces deux concepts présentent la même complexité mais pas le même nombre d'exemples positifs. Dans le contexte d'une tâche de catégorisation, par principe, les sujets doivent séparer les exemples positifs des négatifs. Si on considère à nouveau le concept présenté en Fig. 1b ("triangles et grands"), les exemples positifs sont : tous sauf les "petits ou carrés" (la négation du concept présenté en Fig. 1c), c'est-à-dire : pas "carrés ou petits" ou encore : ni carrés ni petits, donc : "triangles et grands". Cependant, les tâches de catégorisation ne requièrent pas de rappeler les objets dans l'ordre, à la différence des tâches que nous allons proposer. Rappeler six objets dans l'ordre sera plus complexe que d'en rappeler deux, malgré le fait que ces objets présentent un même niveau de complexité (complexité de Feldman = 2 pour "triangles et grands" et pour "petits ou carrés"). Nous avons donc développé une version de la complexité de Feldman pour l'adapter à une présentation séquentielle d'objets.

La complexité algorithmique, quelle que soit la façon dont elle est approchée par le calcul (Feldman, ou complexité algorithmique pour les chaînes courtes), nous permet de proposer des mesures de la mémoire de travail dans lesquelles la chunkabilité du matériel présenté est contrôlée. Nous supposons que les participants ne disposent pas de connaissances en mémoire à long terme concernant le matériel (dont nous développerons plus loin les caractéristiques). Ainsi les regroupements que les participants pourraient faire sont permis uniquement par la structure de l'information présente, en faisant appel à des processus d'abstraction et d'organisation basés sur le traitement "au vol" de l'information présente.

Objectifs de la thèse

Cette thèse a pour objectif de mesurer les capacités de la mémoire de travail à partir de tâches simples de rappel immédiat, dans lesquelles un processus de chunking sera induit. Les séquences manipulées permettront une mesure de leur chunkabilité, en fonction de critères fondés sur la complexité algorithmique. Ces échelles de complexité seront utilisées comme indicateurs de la compressibilité de l'information à mémoriser. Ainsi, caractériser le matériel à mémoriser en terme de compressibilité permet d'émettre l'hypothèse que le processus de chunking repose en partie sur l'optimisation d'un espace de stockage limité en mémoire de travail. Cowan (2016) suggère d'obtenir une capacité en chunks en limitant le traitement articulatoire, afin d'empêcher les regroupements. Dans ce cas les adultes peuvent mémoriser généralement seulement 3 ou 4 chunks (Chen & Cowan, 2009 ; Cowan, 2001 ; Cowan et al., 2012). Nous proposons de mesurer ce même empan tout en permettant aux participants de chunker librement, en maintenant l'hypothèse que quatre est bien la capacité réelle de la mémoire de travail. Ainsi, mesurer un empan de chunks permettrait de rendre compte, dans le même temps, des capacités de stockage de la mémoire immédiate et des capacités de traitement de l'information en question. De ce point

de vue, nous pensons que si nos tâches d'empan peuvent spécifiquement prédire l'intelligence fluide, c'est parce que les capacités de la mémoire de travail permettent de détecter et réorganiser en chunks des régularités dans l'information. L'hypothèse clé relative à la prédiction de l'intelligence naît du lien entre la limite de capacité de la mémoire de travail et la capacité à compresser l'information. Plus on est capable de compresser une information compressible, plus on potentialise la capacité de la mémoire de travail, afin de libérer un espace mental pouvant permettre de tenir des raisonnements plus complexes.

Resumé des articles

Article 1 : The Formation of Chunks in Immediate Memory and its Relation to Data Compression

Cette étude vise à comprendre les processus cognitifs impliqués dans la mémorisation à court terme, et à explorer les stratégies de compression et d'organisation de l'information référant au processus de chunking. Selon les estimations de Miller (1956), la mémoire à court terme est limitée à 7 ± 2 items si le chunking n'est pas contrôlé. Selon les estimations de Cowan (2001), elle avoisinerait plutôt les 4 ± 1 items si le chunking est empêché. Notre objectif était d'évaluer la capacité de la mémoire à court terme, tout en contrôlant les processus de chunking. Notre hypothèse se base sur l'idée que si les sujets utilisent des règles logiques pour réduire l'information, ils doivent regrouper plus facilement deux items successifs s'ils présentent des similarités entre eux et d'autant plus si les items sont organisés de manière à favoriser l'identification de règles logiques permettant de grouper l'information en chunks. Selon notre hypothèse, les indices de complexité (ou de compressibilité) doivent prédire les performances des sujets : plus le matériel présenté est issu d'un concept algorithmiquement simple, plus le rappel doit être performant.

Nous avons construit une tâche de rappel sériel immédiat dans laquelle les participants (67 étudiants) devaient mémoriser et rappeler dans l'ordre et le plus rapidement possible des séries de stimuli images. Il était possible de regrouper ces images en fonction de caractéristiques communes, selon trois dimensions binaires : forme, taille et couleur. Dans cette tâche, la chunkabilité des séquences était modulée par plusieurs facteurs. (1) La longueur : les séquences présentaient entre un et huit stimuli. Pour chaque essai, un ensemble de huit objets était généré par tirage au sort de deux tailles, deux couleurs (parmi huit possibles) et deux formes

(parmi huit possibles). (2) La compressibilité des séquences, prédite par la complexité logique de Feldman (2000), indiquant que des objets peuvent être compressés en une règle simplifiée lorsque l'information peut être réduite. Par exemple, la séquence de deux objets : "grands triangles blancs et petits triangles blancs" peut être simplifiée en "triangles blancs". (3) L'ordre de présentation des stimuli, de manière à favoriser plus ou moins la découverte de règles compressées. Trois conditions d'ordre de présentation ont été retenues sur la base des travaux de Mathy et Feldman (2009) : un ordre basé sur une règle, un ordre basé sur la similarité d'objets successifs et un ordre basé sur leur dissimilarité.

En accord avec nos hypothèses, les résultats ont montré que la longueur de séquence, la complexité de Feldman et le type d'ordre ont influencé significativement le rappel et le temps de réponse. La capacité mesurée à 50% de réponses correctes équivaut à 3.5 items en moyenne, ce qui correspond à la limite évaluée par Cowan (2001). Conformément à nos attentes, plus les séquences présentaient un niveau de complexité élevé, moins le rappel était correct. La complexité de Feldman a donc été un bon indicateur de la compressibilité du matériel présenté. Par ailleurs, le rappel a été plus performant dans la condition dans laquelle l'ordre était fondé sur une règle que dans la condition d'ordre fondé sur une simple similarité, et encore plus que dans la condition dissimilarité. Nos résultats montrent qu'il est envisageable de mesurer la capacité de la mémoire à court terme tout en mesurant les processus de chunking opérant à court terme. Nous en concluons que les individus exploitent directement les régularités présentes dans l'information pour la compresser très rapidement en mémoire immédiate.

Article 2 : There is no fundamental difference between simple and complex span tasks

Il est établi que le processus de chunking (utilisant des informations en mémoire à long terme) peut faciliter le rappel immédiat. Nous avons montré dans l'article 1 que le recodage de séquences d'items peut se produire très rapidement en mémoire immédiate sur la base d'information compressible au sein même d'une séquence présentée. Dans ce second article, nous montrons l'intérêt d'utiliser des séquences, a contrario, non compressibles, c'est à dire pour lesquelles le chunking est peu probable. Nous avons donc construit une première expérience dans laquelle plusieurs conditions sont confrontées, en croisant deux facteurs, d'une part la chunkabilité du matériel et d'autre part la présence ou non d'une tâche concurrente. Concernant la compressibilité des séquences, soit elles présentaient des régularités compressibles ou "chunkable", soit les séquences étaient plus hétérogènes, "non chunkables", c'est à dire offrant pas ou peu de possibilités de regroupement logique. En conséquence, quatre conditions ont été construites : une tâche d'empan simple, utilisant du matériel chunkable ; une tâche d'empan complexe, utilisant du matériel chunkable ; une tâche d'empan simple, présentant du matériel non-chunkable ; et enfin une tâche d'empan complexe, utilisant du matériel non-chunkable. Notre prédiction était que le chunking serait possible dans le cas des tâches simples, mais qu'il ne pourrait pas intervenir lorsque l'attention est dévolue à une tâche concurrente dans la tâche d'empan complexe. Cela signifie que seules les tâches d'empan simples peuvent être sensibles à la présence de matériel chunkable.

Dans une deuxième partie, nous proposons quatre expérimentations utilisant des tâches d'empan simples dans lesquelles les participants devaient mémoriser des séquences d'objets ne présentant pas de régularités afin de mesurer des capacités de la mémoire immédiate sans qu'un processus de chunking n'intervienne. Nous faisons la seconde hypothèse que bien que la

tâche soit simple, le matériel rendu non compressible devrait produire une limite de capacité autour de 3 ou 4, tout comme dans les tâches d'empan complexes classiques utilisant du matériel d'empan simple.

Pour nuancer notre première hypothèse, nous avons constaté que les tâches d'empan simples et complexes étaient affectées de façon relativement similaire par la présence de chunks. Par conséquent, nous concluons que les deux types d'empan ne mesurent probablement pas des concepts théoriques aussi différents, tout du moins en ce qui concerne l'implication de la composante de traitement dans le stockage. Nous montrons néanmoins, en parfait accord avec la seconde hypothèse, que notre matériel non compressible offre une estimation résolument basse des capacités en mémoire de travail.

Article 3 : Chunking in Working Memory and its Relationship to Intelligence

Au cours d'une mesure d'empan, on demande habituellement au participant de rappeler simplement une liste de chiffres, de lettres ou de mots. Cependant, s'ils ne sont pas contrôlés, des processus tels que la réorganisation de l'information en chunks et le recours aux connaissances présentes en mémoire à long terme, peuvent intervenir pour augmenter l'empan. Notre objectif était d'estimer plus finement l'implication en mémoire immédiate des processus de chunking au vol, c'est à dire basés sur l'information présente, sans recours à des informations en mémoire à long terme. Pour cela, nous avons développé des nouvelles tâches d'empan simples permettant une mesure de complexité algorithmique récente. Cette mesure de complexité permet d'évaluer la compressibilité de l'information, donc par analogie les possibilités de chunking. Nous pensons qu'il existe un lien entre la capacité des participants à chunker l'information dans ces tâches et leur performance aux matrices de Raven, un test d'intelligence fluide, car toutes deux solli-

citent une optimisation de l'information (l'optimisation de l'espace libère des ressources pour développer des tâches de raisonnement plus complexes). Dans cet article, nous comparons entre elles les performances des participants aux matrices de Raven, à des tâches d'empan classiques (empan simple et complexe) et enfin, aux tâches que nous avons développées.

Dans deux expériences, nous avons montré que les tâches d'empan impliquant un processus de chunking prédisaient mieux les performances aux matrices de Raven que d'autres tâches d'empan complexes ou simples. Cette étude conclut que le traitement et le stockage en mémoire de travail devraient être examinés conjointement, particulièrement lorsque le traitement est entièrement dédié aux éléments stockés. Nous concluons que l'interaction entre le stockage et le traitement (permettant à un processus de chunking d'opérer en mémoire immédiate) fournit un indice novateur de la capacité en mémoire de travail. Ceci est en accord avec Unsworth et al. (2009) qui soutiennent que le traitement et le stockage devraient être examinés ensemble parce que la mémoire de travail est capable de traiter et de stocker des informations en même temps. L'originalité de notre approche est néanmoins de nous distinguer de la majorité des études utilisant exclusivement des tâches complexes. Si ces tâches complexes sollicitent bien traitement et stockage, elles ne permettent pas d'étudier les performances en mémoire lorsque le traitement est au service du stockage.

Nota bene : les articles suivants ont été compilés pour être intégrés de façon continue dans la thèse, ainsi, les numéros de figures, de tableaux et de notes de bas de page sont continus et ne correspondent pas à la numérotation originale des articles.

Deuxième partie

Formation de chunks en mémoire immédiate et compression de données

The Formation of Chunks in Immediate Memory and its Relation to Data Compression

Mustapha Chekaf

Université de Franche-Comté

Nelson Cowan

University of Missouri-Columbia

Fabien Mathy [1]

Université Nice Sophia Antipolis

1. This research was supported by a grant from the Agence Nationale de la Recherche (Grant # ANR-09-JCJC-0131-01) awarded to Fabien Mathy, a grant from the Région de Franche-Comté AAP2013 awarded to Fabien Mathy and Mustapha Chekaf, and by NIH Grant R01 HD-21338 awarded to Nelson Cowan. We are grateful to Jacob Feldman, Ori Friedman, Alessandro Guida, Harry H. Haladjian, and the Attention & Working Memory Lab at Georgia Tech for their helpful remarks.

Abstract

This paper attempts to evaluate the capacity of immediate memory to cope with new situations in relation to the compressibility of information likely to allow the formation of chunks. We constructed a task in which untrained participants had to immediately recall sequences of stimuli with possible associations between them. Compressibility of information was used to measure the chunkability of each sequence on a single trial. Compressibility refers to the recoding of information in a more compact representation. Although compressibility has almost exclusively been used to study long-term memory, our theory suggests that a compression process relying on redundancies within the structure of the list materials can occur very rapidly in immediate memory. The results indicated a span of about three items when the list had no structure, but increased linearly as structure was added. The amount of information retained in immediate memory was maximal for the most compressible sequences, particularly when information was ordered in a way that facilitated the compression process. We discuss the role of immediate memory in the rapid formation of chunks made up of new associations that did not already exist in long-term memory, and we conclude that immediate memory is the starting place for the reorganization of information.

Chunk Formation in Immediate Memory

and How It Relates to Data Compression

Introduction

Individuals have a tendency to make information easier to retain by recoding it into chunks (e.g., Cowan et al., 2004). The process of chunking simplifies memorization by taking advantage of knowledge to reduce the quantity of information to be retained (Miller, 1956). As a key learning mechanism, chunking (or grouping) has had considerable impact on the study of expertise (e.g., Chase & Simon, 1973 ; Ericsson et al., 1980 ; Hu & Ericsson, 2012), immediate recall (e.g., Chen & Cowan, 2005 ; Farrell, Wise, & Lelièvre, 2011), and memory development (e.g., Cowan et al., 2010 ; Gilchrist, Cowan, & Naveh-Benjamin, 2009).

For chunking to benefit memory, people need to be able to retrieve the chunks they stored. One way people retrieve chunks is via long-term memory processes (French et al., 2011 ; Gobet et al., 2001 ; Guida et al., 2012 ; Reder, Liu, Keinath, & Popov, 2016). Consider the letter string IBMCIAFBI. As Miller discussed, this letter string can be easily simplified to form three chunks if one uses long-term memory to recall the U.S. agencies (Miller, 1956) whose acronyms are IBM, CIA, and FBI.

Previous work on chunking has focused on how long-term memory aids chunk creation. However, immediate memory might also play a fundamental role in the creation of chunks. People may form chunks in immediate memory by rapidly encoding patterns before any consolidation in long-term memory occurs. For example, it is easy to remember the letter string AQAQAQ using a simple rule of repetition (e.g., AQ three times). This type of simplification does not necessarily depend on the use of long-term memory to recall past knowledge that relates items

to each other [2]. Instead, this process depends on the apprehension of regularities inherent to the stimulus at hand, i.e., compression.

This idea that immediate memory might play a fundamental role in the creation of chunks has generally been overlooked. Some previous findings are consistent with the proposal that chunks can increase memory capacity (Brady et al., 2009 ; Feigenson & Halberda, 2008). However, these studies have mostly focused on how long-term-memory representations contribute to encoding in immediate memory. In contrast, our goal is to provide a principled quantitative approach to how immediate memory relates to the formation of chunks. Getting a larger picture of chunking as a process originating in immediate memory needs a precise conceptualization, and the concept of compressibility could help in doing so.

We propose a two-factor theory of the formation of chunks in immediate memory. The first factor is compressibility (i.e., the idea that a more compact representation can be used to recode information in a lossless fashion [3]). Compressibility could predict chunking because it measures the degree to which the material is patterned, and hence the degree to which the material can be simplified. Memory for compressible sequences should be superior to memory for non-compressible sequences (the same way that studies in the domain of categorization have shown that compressible material is better learned over the long term; see Feldman, 2000).

The second factor is the order of the information to memorize. Presentation order might influence the ease with which patterns or regularities in the stimuli can be discovered, and compression algorithms typically depend on this kind of information. A presentation order that aligns

2. Long-term memory is needed to retrieve the individual items (e.g., A, Q, ×, and 3), but it is not needed to retrieve combinations of them.

3. By this, we mean a compression process without loss of information (the original data can be accurately reconstructed from the compressed data), and not a "lossy" form of compression (which brings to mind many of the applications in information technology used today to achieve a more substantial reduction of data); see Li & Vitányi, 1997.

with the process of simplifying the material may increase the likelihood that chunking occurs. In contrast, presentation orders that do not aid in the discovery of regularities, might result in failure to chunk compressible materials, causing them to be remembered in a way similar to non-compressible materials. Presentation order should therefore interact with compressibility. As a simple example, one can compress the set "2, 3, 4, 5, 6" with the rule, "all numbers between 2 and 6", whereas with the series "2, 4, 6, 3, 5", that same rule might not be noticed by the participant, so compression might not take place.

This two-factor theory is adapted from the domain of categorization, which has provided a framework for studying category formation in long-term memory, with explanations based on the compressibility of descriptions (Bradmetz & Mathy, 2008 ; Feldman, 2000, 2003 ; Goodwin & Johnson-Laird, 2013 ; Lafond, Lacouture, & Mineau, 2007 ; Vigo, 2006) using different types of presentation orders (based on rules, similarity, or dissimilarity; see Elio & Anderson, 1981, 1984 ; Gagné, 1950 ; Mathy & Feldman, 2009 ; Medin & Bettger, 1994). This framework nicely accounts for a wide range of categorization performance in long-term memory, but could in principle provide similar predictions for immediate memory. Our theory is that a compression model (e.g., Feldman, 2000) can be adapted to immediate memory. The rationale is that elementary structures, i.e., the redundancies that make a structure compressible, are simple enough to be used rapidly in immediate memory to cope with new situations.

We conducted an experiment to test the proposal outlined above, namely, that chunk formation occurs in immediate memory to optimize capacity before any consolidation process in long-term memory occurs. Our prediction is that immediate-memory span is proportional to stimulus compressibility, but only when the order of the information allows the participant to spontaneously detect redundancies such as pairs of similar features.

In the Discussion, we provide ample evidence that there are two major classes of concurrent models that cannot provide correct predictions for our results. The first class is Interference-based models of short-term memory, which predict poorer performance when participants see sequences containing similar features, whereas our model predicts that participants can take advantage of these similarities to compress information. The second class includes the minimal description length (MDL) approaches to long-term memory, which rely on the repetition of trials, and as such, offer no predictions about the compression process at play in our task.

Method

Two key aspects were investigated in the present experiment: compressibility of a sequence and presentation order within a sequence. These two factors were studied using categorizable multi-dimensional objects, with discrete features, such as small green spiral, large green spiral, small red square. The sequences used could not conform to already-learned chunks. Although the features themselves are part of basic knowledge, we are reasonably confident, for instance, that none of our participants had the exact sequence of items "a small green spiral followed by a large purple pentagon and a small yellow pentagon" in long-term memory before starting our experiment. The procedure used a serial recall task, which allowed to study the incremental encoding of chunks. The duration of the display of the memory items and the number of memory items were two other manipulated factors we thought would help us look into the incremental encoding of the chunks.

Participants

Sixty-seven students enrolled at the University of Franche-Comté, $M = 22$ years old ($sd = 2.7$), volunteered to participate in the experiment.

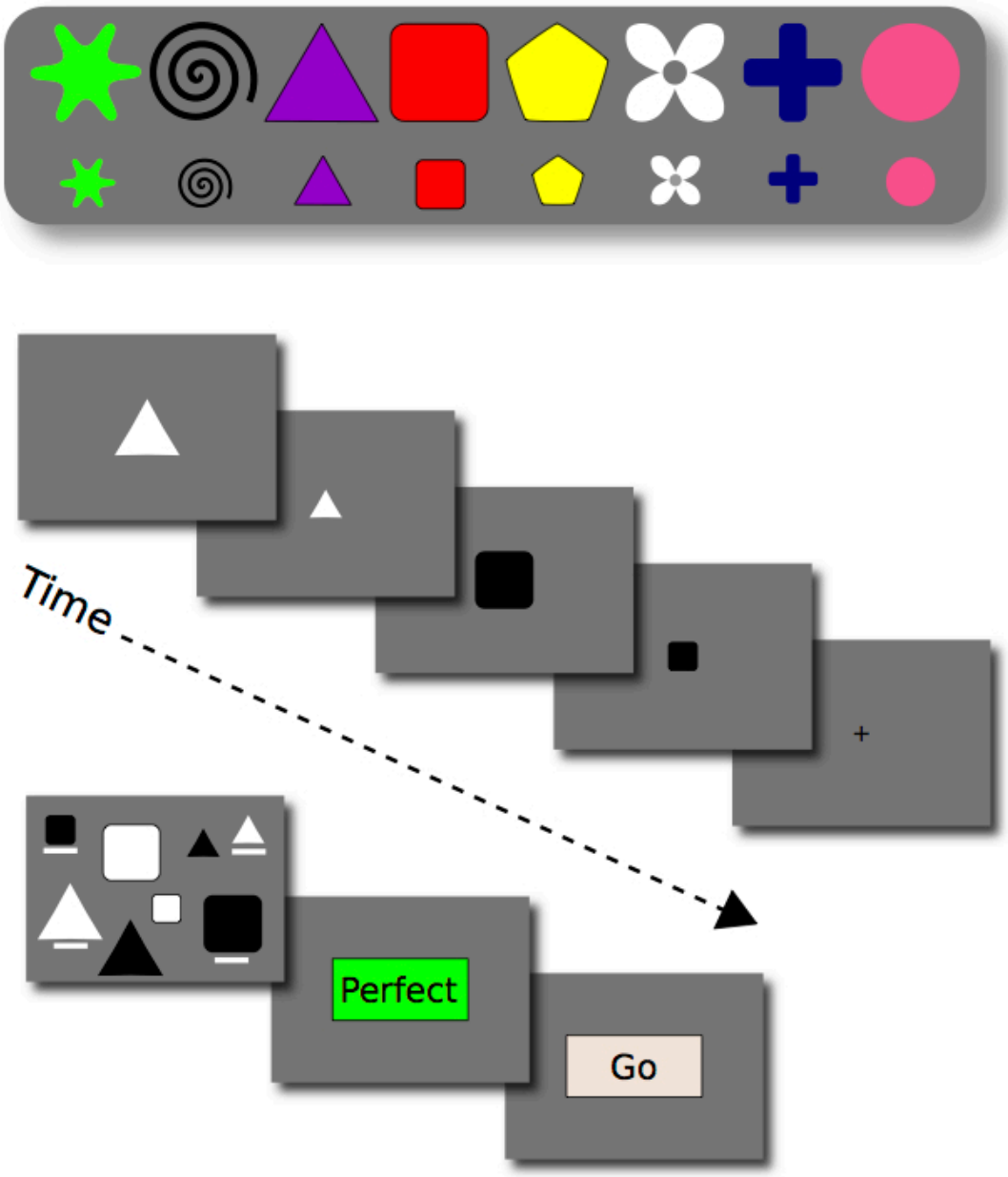


Figure 2 – Screen shots showing a sample of possible stimuli (top panel), a to-be-recalled patterned sequence of stimuli (middle panel), and an example of a response screen that asked the participant to recognize the previous stimuli and to order them by clicking on a mouse (bottom panel).

Stimuli

Our stimuli could vary according to three dimensions: shape, size and color. A combination of two shapes, colors, and sizes makes a set of eight different objects. There were eight different values for the shape dimension and the color dimension (Fig. 2, top panel). However, we restricted the size dimension to two values (large *vs.* small, or 280 × 280 pixels *vs.* 140 × 140 pixels). Shape, size and color are typically used by category learning researchers to build canonical stimulus sets because these dimensions can be easily and clearly partitioned.

For a given sequence, the program randomly chose two out of eight shapes and two out of eight colors (see Fig. 2, top panel), in order to create a set of eight objects. For example, if the values "triangle", "square", "white", and "black" were drawn, the program generated $2 \times 2 \times 2 = 8$ stimuli by combining three features for each stimulus (e.g., small white triangle, large white triangle, ..., large black square). These values allowed for 1568 possible sets of eight objects, so that the probability of a participant coming across two identical sets during the experiment would be very low. The stimuli were presented against a gray background.

Categories

We selected different categories of objects, which were to be displayed and recalled serially. An example is the sequence □ ■ □ ■ △ ▲, which can be represented by six individual exemplars (i.e., large white square, large black square, small white square, small black square, small white triangle, and small black triangle). Following Feldman (2000), this sequence can be redescribed accurately by a shorter logical rule provided that order does not matter ('squares or small', using inclusive disjunction, or 'not[large and triangle]' using conjunction, which by de Morgan's law are equivalent). Another example is the sequence ■ ▲ □ △ ('small black square, small black

triangle, small white square, small white triangle'), which can be simplified by abstracting the feature common to the four objects: "small". Hence, the information for this category is even more compressible and does not require much mental effort to be retained. The to-be-recalled categories of stimuli were chosen on the basis of the exhaustive list of the 21 categories shown in the Appendix.

Ordering of the categories in the sequences

Our measure of complexity only serves here to predict the chunkability of the category set, not the memorization of the sequence. Controlled presentation orders were thus used to manipulate category compressibility. This second manipulation consisted of ordering the objects according to three main types of orders previously developed in the categorization research: rule, similarity, and dissimilarity.

In the **Rule condition** (Fig. 3), the objects were grouped by clusters (sub-categories). The differences were minimal within the clusters, but more marked between clusters (whenever clusters could be found). A minimal degree of similarity between clusters contributes to greater discrimination of the clusters into distinguishable units, thus favoring the discovery of a logical rule. Figure 3 shows the example of an organized sequence in the Rule condition. The shortest rule for describing the set of objects is "square or small". In Figure 3, the first cluster is made up of the large white square (first item) and the large black square (second item), which differ by only one feature. The small white square (third item) that follows the large black square (second item) marks a separation between the first and second clusters because the two contiguous items differ by two features. In the Rule condition, such leap can facilitate the identification of clusters and therefore might induce the formation of chunks based on the clusters (see Fig. 3, bottom, which shows two leaps as a function of time, one after the

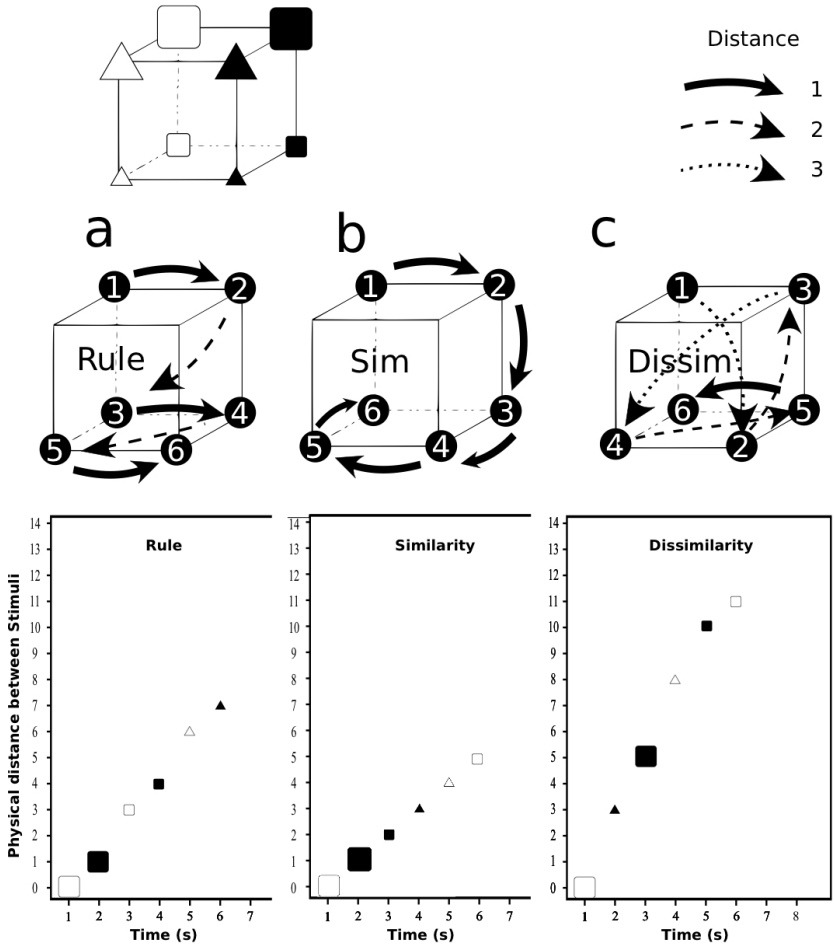


Figure 3 – Example of three sequences from the category structure "square or small", representing six items (top left cube). *Note.* The six items are replaced by black numbered circles on each of the three cubes to simulate their sequencing. In each of the three cubes, the presentation order is indicated by the numbers from one to six and by the arrows. The distance between two consecutive objects is described by the type of arrow: solid (one edge, or one feature difference), dashed (two edges), or dotted (three edges). In the Rule condition, the objects are presented in three clusters, within which the solid arrows are parallel and go in the same direction. The regularity of the rule is related to two factors: the separation of the clusters (dashed arrows) and the similarity between the objects in the same cluster (solid arrows), both of which facilitate the formation of small groups. In the Similarity condition, the inter-stimulus distance is minimal. All the objects are linked to each other by solid arrows, which can potentially make a unique chunk of six objects. In the Dissimilarity condition, the sequence is characterized by a maximal inter-stimulus distance. The three plots at the bottom show the distances between the stimuli as a function of presentation time (1 second per item). For instance, the first plot shows two distance leaps in the Rule condition, which can facilitate the identification of three chunks. The leaps are more numerous, larger, and less regular in the Dissimilarity condition, while there are no leaps at all in the Similarity condition.

second item and the other after the fourth item). In this condition, the participants should have the best recall performance, particularly if they attempt to form small groups of objects by compressing information locally [4].

In the **Similarity condition** (Fig. 3), a sequence was chosen to favor a minimal overall (i.e., for the entire sequence) inter-stimulus distance. The objects were then presented as a string following the principle that there was maximal similarity between two successively presented objects. Figure 3 shows an example of a similarity order for the same category of objects. Notice that adopting a similarity-based grouping process for this sequence would result in retaining a single long chunk of six items.

Finally, in the **Dissimilarity condition** (Fig. 3), the set of objects was chosen so as to favor a maximal total inter-stimulus distance. Each object presented minimal similarity with the preceding one. This condition deliberately disorganizes the presentation to make the associations between stimuli more difficult, and should hinder the chunking process. Using a dissimilarity-based strategy would result in retaining several separate small chunks of independent items. Consequently, recall performance should be lower than in the other two conditions. Figure 3 shows an ordered sequence in the dissimilarity condition using the same example as above.

The 21 categories of objects shown in Fig. 6 in the Appendix were transformed into 51 sequences, according to these three presentation orders. The number 51 results from the fact that the order of six categories could not be manipulated. This was the case for four categories made up of one or two objects (once the first object is drawn, only the second object remains), and for two categories, for which the distances are identical between any two objects in a pair (this

4. This rule-based presentation order is a bit different from the one used by Mathy and Feldman (2009), in which the order of the objects within the clusters was random. Because our task is serial, our choice here consisted of ordering the objects logically within the clusters to encourage participants to connect the objects serially using the shortest description.

is the case for the category in the left column of Fig. 6, in which FC (Feldman's measure of complexity) = 8 and FC = 10) [5]. These six categories were coded "order condition = None". The Rule, Similarity and Dissimilarity conditions were applied to the 15 remaining categories. Each participant saw all 51 sequences ($3 \times 15 + 6$).

Procedure

For each participant, the 51 sequences were presented in random order to avoid ascending list length or complexity. We established six possible rotations for each category structure that corresponded to six possible ways to place the objects on the diagram while keeping the same structure. One rotation (among the six) was randomly drawn for each sequence presented so as to multiply the possible combinations in dimensional terms (shapes, sizes and colors). Thus, the participant never knew in advance which dimension would be the most relevant for the categorization process. For the experiment as a whole, 3417 sequences/trials (51 $\times$ 67 participants) were presented.

Serial report was investigated using a procedure similar to that used in the visual STM serial report task (Avons & Mason, 1999 ; Smyth, Hay, Hitch, & Horton, 2005), except that the stimuli were mixed with distractors on the response screen. Depending on the sequence, one to eight stimuli were displayed serially in the center of the screen (see Fig. 2, middle panel) at intervals of one second (41 participants) or two seconds (26 participants) per stimulus, depending on

5. In fact, the most incompressible categories provided fewer possibilities for order manipulation. For instance, the concept △□ ▲ (i.e., *white [small triangle or large square] or large black triangle*, in the 7^{th} cell in the left column of Figure 6) was typical of a complex concept for which different presentation orders could not be specified. The reason is that the objects in every pair of stimuli have a single feature in common. In this particular case, once a first object is chosen, there is no other choice than to pick a second object that is two-features away from the first, and so on. As a result, the rule-based/similarity-based/dissimilarity-based distinction is no longer relevant. However, because the chosen order cannot benefit from any simple logical rule, or any simple similarity-based relationship, the order can be considered dissimilarity-like.

the condition [6]. During the recall phase (see Fig. 2, bottom panel), the original set of eight stimuli was displayed randomly on the screen. The stimuli were underlined (using a white line) when the user clicked on them. After the user validated his/her answer with the space bar, a feedback screen indicated if the recall was correct (i.e., both item memory and order memory had to be correct), and then a screen with a GO window appeared and the user moved on to the following sequence by pressing on the space bar. The experiment lasted an average of 25 minutes.

Results

The analyses were conducted on correct (1) or incorrect (0) serial-recall scores for each trial (a response was scored correct when both the items and the positions were correctly recalled), and for the average recall score across conditions (proportion correct). The data was first aggregated for a given variable, e.g., 'presentation time', in order to run a separate univariate ANOVA for each dependent variable (e.g., the mean proportion correct for all trials pooled).

Summary of the expected results

(1) Recall performance should depend on sequence length to the extent that longer sequences require a representation with more chunks. (2) A higher compressibility of information (i.e., lower FC within sequences of the same length) should result in better recall because higher compressibility enables better recoding of the entire set of items. (3) A greater degree of regularity

6. The 1 s *vs.* 2 s duration was a between-subject factor. The one-second condition tends to be the standard condition for measuring memory span, so we used this condition with a large number of participants. However, we still wanted to modulate the time allotted to chunking by extending the display duration (i.e., 2 s) with a smaller group of participants, because display duration can affect memorization strategies (Hockey, 1973). The two-second condition was run with only 26 participants to make sure the participants in the more standard one-second condition had sufficient time to parse the stimuli and achieve similar performance.

in the presentation order (rule-based, followed by similarity-based, followed by dissimilarity-based) should favor the compression of the available regularities into newly formed chunks. (4) For the less important factor, display duration, we expected better recall performance for the longer duration.

Proportion correct

Regarding the 2 s *vs.* 1 s conditions, a preliminary between-subject analysis on proportion correct (averaged across trials) showed that stimulus duration was significant ($t(65) = -2.5, p = .014$, $\eta^2 = .09$), with respective means equal to .33 ($sd = .11$) and .27 ($sd = .08$). However, because recall was only slightly higher in the 2 s condition, with a rather small size effect, we chose not to keep the 1 s *vs.* 2 s factor as a moderator in the subsequent analyses.

Figure 4 shows the mean proportion of correct responses as a function of the length of the sequence presented, by the type of presentation order. Overall, a nonlinear regression (using an s-shaped sigmoid function of the form $a - 1/(1 + exp(-b \times x + c))$ of the mean points on this figure (along the x-axis) showed that 78% of the performance variance was explained by sequence length ($R^2 = .78$). Regardless of order, the memory-span estimate at a threshold of 50% correct responses was 3.5 objects. When the length of the sequence increased from three to four objects, mean performance dropped from .64 to .33. It also diminished by half again as much between four and five items to recall (from .33 to .14). This impressive drop resulted in performance at the lower end of Cowan's (2001) memory-span estimate of 3-5 items, and is similar to other more specific estimates of 3-4 items (Broadbent, 1975 ; Chen & Cowan, 2009 ; Luck & Vogel, 1997).

However, a multiple linear regression analysis on mean proportion correct showed that each of the three factors contributed significantly to the drop in performance. The percentages

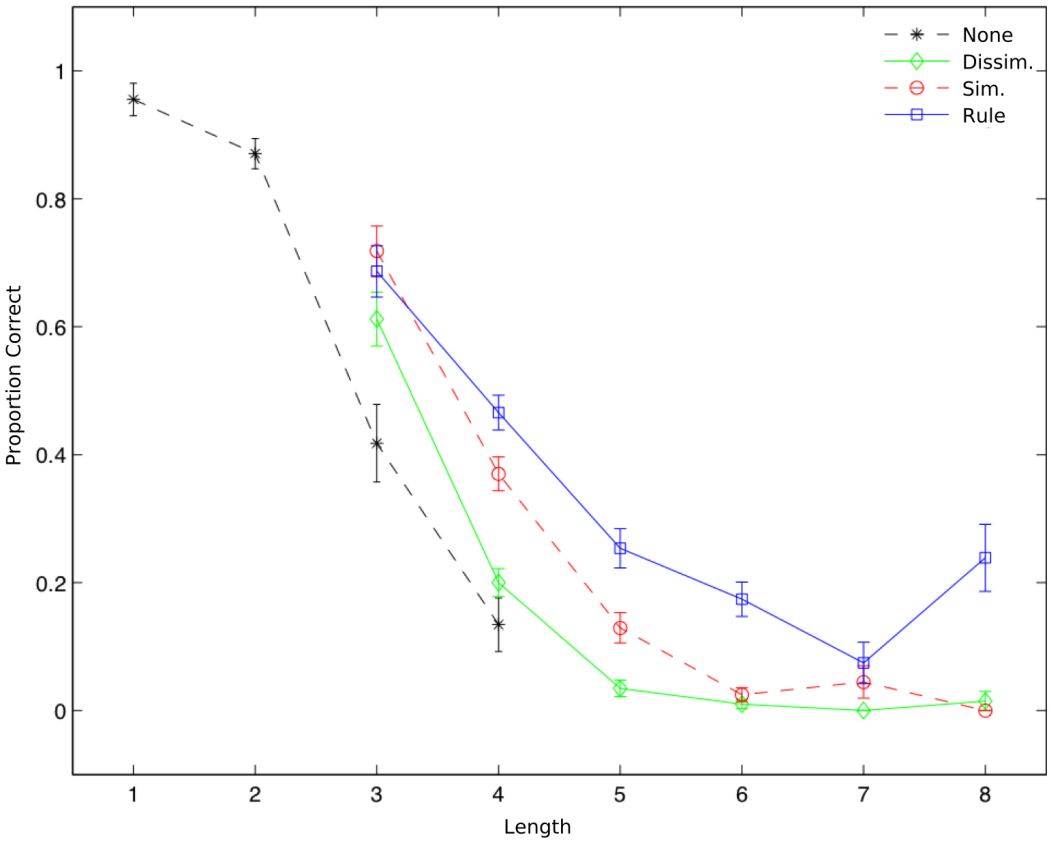


Figure 4 – Proportion of correct sequences recalled as a function of the sequence length (number of objects), for each presentation order. *Note*. The error bars show ±1 standard error.

of variance explained were as follows: sequence length ($\beta_i r_{Yi} = 67\%$), compressibility/FC ($\beta_i r_{Yi} = 9\%$, which included eight unique values, that is 1, 2, 3, 4, 5, 6, 8, 10), presentation order [7] ($\beta_i r_{Yi} = 12\%$), and presentation time ($\beta_i r_{Yi} = 1\%$), totaling 89%, $F(6, 89) = 110, p < .001, R^2 = .89$. The effects of length and presentation order on performance are clearly visible in Figure 4, whereas the effect of compressibility is shown in Figure 5. Regarding the effect of FC, a partial correlation analysis showed that at a constant sequence length, compressibility and correct proportion were negatively correlated ($r = -.28, p < .001$). Indeed, the more complex the sequence, the lower the performance (see Fig. 5, left). The ANOVA with repeated measures, with FC as a within-subject factor and proportion correct as the dependent variable yielded a significant effect of FC for most sequence lengths: $F(2, 132) = 2.9$, $p = .054$, $\eta_p^2 = .04$ for two-object sequences; $F(2, 132) = 11.5$, $p < .001$, $\eta_p^2 = .15$ for three-object sequences; $F(3, 198) = 45.6$, $p < .001$, $\eta_p^2 = .41$ for four-object sequences; $F(2, 132) = 18.8$, $p < .001$, $\eta_p^2 = .22$ for five-object sequences; and $F(2, 132) = 45.1$, $p < .001$, $\eta_p^2 = .41$ for six-object sequences. Figure 5 shows the linear trend that we obtained using data collapsed across participants, complexity, and sequence length.

The proportions of correct recall for the Rule, Similarity, and Dissimilarity orders were respectively, .35 (sd = .16), .25 (sd = .13), and .16 (sd = .09). The ANOVA with repeated measures yielded a significant effect, $F(2, 132) = 80, p < .001, \eta_p^2 = .55$. We also observed significant differences between the Rule and Similarity conditions, $t(66) = 6.8, p < .001$, and between the Similarity and Dissimilarity conditions, $t(66) = 7.8, p < .001$, even after Bonferroni correction.

At lengths 3 and 4 (the conditions for which the four types of orders applied), the reason for

7. Because the use of dummy variables is more appropriate for representing multiple groups (Cohen & Cohen, 1983), the Rule, Similarity, Dissimilarity, and None conditions were recoded using dummy variables.

the lower score in the None order is linked to the complexity of the heterogeneous categories, for which FC = 8 and FC = 10. In these conditions, the proportions of correct recall for the Rule, Similarity, Dissimilarity, and None orders were respectively, .52 (sd = .03), .47 (sd = .03), .32 (sd = .02), and .28 (sd = .04). The ANOVA with repeated measures yielded a significant effect, $F(3, 198) = 25, p < .001, \eta_p^2 = .28$. We also observed significant differences between the Rule and Similarity conditions, $t(66) = 2.11, p = .04$, and between the Similarity and Dissimilarity conditions, $t(66) = 6.7, p < .001$, but the difference between the Dissimilarity and None conditions was not significant. At lengths 3 and 4, the only pair left significant after Bonferroni correction was Similarity-Dissimilarity.

A repeated-measures ANOVA was run on presentation order (with the rule-based, similarity, and dissimilarity levels) and FC (with levels 1, 2, 3, 4, 5, 6, and 8; note that the value 10 could not interact with presentation order, and also, the N/A value could not be included to test a linear trend). Feldman's complexity significantly affected the mean proportion correct, $F(6, 396) = 76, p < .001, \eta_p^2 = .54$, and we found a linear trend for this factor, $F(1, 66) = 235, p < .001, \eta_p^2 = .78$. We also found a significant interaction between the two factors, $F(12, 792) = 8.5, p < .001, \eta_p^2 = .11$, which tended to show that the Rule-order benefited the participants more when FC was low.

Transitional error probabilities

Transitional error probabilities (N. F. Johnson, 1969a, 1969b) can provide further evidence as to whether the participants actually used the available chunks. Without such an analysis, there is no clear evidence of what chunks were adopted during the task. Based on the Johnson's analyses, one can explore the relationship between performance on two adjacent items when they can be assumed to fall within the same chunk, rather than on opposite sides of a chunk

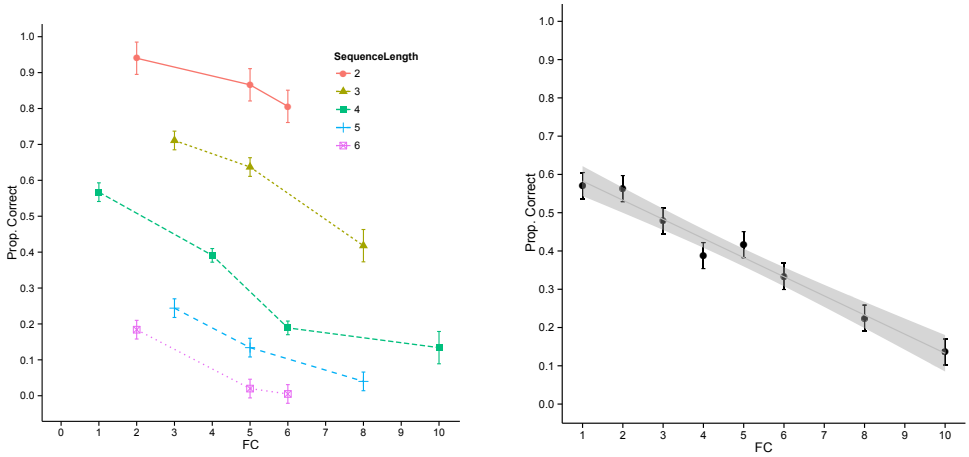


Figure 5 – Proportion of correct sequences recalled as a function of Feldman's Complexity (FC), across each sequence length (left plot) and collapsed (right plot). *Note.* The analysis shown in the right plot is based on the 3217 trials, to include the entire variance. The error bars show ± one standard error.

boundary. For instance, the sequence □■■□ could be chunked into □■ and ■□ using a logical rule. This sequence thus contains two potential within-chunk transitions (e.g., □ to ■ and ■ to □) and a single between-chunk transition (■ to ■). The prediction was that within-chunk transitions would more often be made correctly than between-chunk transitions would, under the assumption that participants use the available chunks and store them under the same memory code.

For all adjacent items, a transition was scored correct when the second element was recalled, provided the first item had been correctly recalled. A transition error was counted whenever the second element was not recalled but the first item had been correctly recalled. If the participant's response was □■■, the analysis indicated that there was only one out of two correct within-chunk transitions and that the single between-chunk transition was correctly made. A transition was not counted when the first item of an adjacent pair was not recalled.

The results reported in Table 1 show that a greater proportion of transitions were correctly

made within chunks than between chunks, and that this difference was particularly great for the rule-based presentation order. Another expected result was that recall decreased more for within-chunk pairs than for between-chunk pairs across all orders. The reason is that rule- and similarity-based orders are assumed to facilitate chunk encoding more. Table 1 shows this interaction between presentation order and the two types of pairs. A repeated-measures ANOVA using order (rule-based, similarity-based, dissimilarity-based) and type of pair (within-chunk *vs.* between-chunk) as two within-subject factors confirmed this interaction, $F(2, 132) = 7.2, p < .001, \eta_p^2 = .10$, as well as a main effect of these two factors ($F(2, 132) = 158, p < .001, \eta_p^2 = .71$, and $F(1, 66) = 326, p < .001, \eta_p^2 = .83$, respectively).

Note that the above analyses argue in favor of the idea that chunk identification was rule-based. To better account for the formation of chunks in the rule-based order and the similarity-based order, a more precise and neutral description of how chunks might be incrementally encoded during the task is presented below

Incremental encoding of chunks of varying lengths using strict position scoring

Finally, we look at how much information participants may have encoded irrespective of whether or not the response was perfectly recalled. To evaluate the actual number of items encoded, we present a microanalytical characterization of how chunks can be encoded incrementally during the task. First, we identified the chunks that could theoretically be formed by binding two or more adjacent objects based on their similarity. This new analysis is more neutral than the previous ones in that it does not favor the rule-based process. We posited that the probability of forming a chunk was maximal when two adjacent objects differed by only one feature. Each study sequence was decomposed into chunks when at least two adjacent objects were considered similar. A chunk could, however, contain three objects if the third object also differed by only

one feature from the second object, and so on, when more than three objects could be strung together. Put differently, a new chunk was formed whenever an object differed by at least two features from the preceding object. A single-object chunk was formed whenever there was no way of grouping two objects together. Using this method, chunk size therefore ranged between one object long and eight objects long.

For instance, the following study sequence □■■▲▲□ was split as follows to form a new chunked sequence: □■■–▲▲–□ , using three chunks of varying lengths (3, 2, 1, respectively) and totaling six encoded objects. The ratio of six objects to three chunks (i.e., two objects per chunk on average) indicates the compression achieved by the recoding process.

To find the number of chunks that the participants encoded, their response sequences were split using the same encoding process before being aligned to the chunked study sequence. A partial-credit scoring system was used to get the sum of the chunks correctly recalled in their correct positions. For instance, had the participant recalled □■■▲□ , this response sequence was split into □■■–▲–□ before being aligned to the chunked study sequence □■■–▲▲–□. The alignment would only score the first and third chunks as correctly recalled in their correct positions. The dissimilarity-order condition was removed from this analysis because, by construction, it generally contained one-object chunks. Across all sequence lengths, the theoretical average of to-be-encoded chunks in the rule-based order was 2.9 *vs.* 1.5 in the similarity-based order. Conversely, the number of objects that could be unpacked from one chunk had a theoretical average that was greater in the similarity-based condition (4.0) than in the rule-based condition (1.7). By construction, the similarity-based condition favored longer chunks (and consequently fewer chunks), whereas the rule-based condition rarely offered the opportunity to form chunks made of more than two objects.

Table 1 – *Average proportion (all participants pooled) of adjacent two-stimulus sequences recalled, as a function of presentation order and pair type.*

Type of pair	Rule-based	Similarity-based	Dissimilarity-based
Within-chunk	.56 (.14)	.45 (.13)	.29 (.11)
Between-chunk	.43 (.15)	.35 (.13)	.21 (.09)

Note. Parentheses indicate SDs.

The results indicated a larger number of chunks encoded in the rule-based order ($M = 1.28, sd = .42$) than in the similarity-based order ($M = .45, sd = .20$), after averaging the participants' performance across sequence lengths ($t(66) = 20.2, p < .001, \eta_p^2 = .86$). The participants recalled an average of 2.27 objects in the rule-based order versus 1.13 objects in the similarity-based order, and this difference reached the same significance level. The ratio between these two counts indicates that the participants encoded an average of $2.27/1.28 = 1.77$ objects per chunk in the rule-based order and $1.13/.45 = 2.5$ objects in the similarity-based order. This difference was also significant after averaging participants' performance across sequence lengths ($t(66) = 13.7, p < .001, \eta_p^2 = .74$). In comparison to the theoretical number of objects that could be encoded in chunks in the rule-based condition (1.7, as indicated above), the participants could potentially encode these chunks perfectly (1.77 objects per chunk), while retaining fewer chunks than available (1.28 instead of 2.9). Information was therefore correctly encoded into in chunks but the number of chunks was subject to a capacity limit.

Incremental encoding of chunks of varying lengths, regardless of position

We used a less stringent free-position scoring system that counted a chunk as correctly recalled no matter where it was in the response sequence. We then determined the number of objects that could be unpacked from the recalled chunks. Using this method, the number of objects unpacked from the chunks found in the participants' responses was significantly greater in the rule-based condition ($M = 2.99, sd = .65$) than it was when strict position scoring was used

($M = 2.27, sd = .74$), $t(66) = 18.4, p < .001, \eta_p^2 = .84$, and as such, a more plausible estimation of the expected 4±1 span was attained. Recall in the similarity-based condition was also slightly greater, increasing from 1.13 to 1.34 ($t(66) = 12.4, p < .001, \eta_p^2 = .70$).

Incremental encoding of chunks corresponding to pairs of items in the similarity-based condition

We carried out a final test to see whether participants in the similarity-based condition encoded smaller chunks than those available, using an encoding process similar to the one induced in the rule-based condition (i.e., encoding objects by pairs). For instance, the sequence □■■□ is potentially a single four-object chunk since none of the transitions between adjacent objects involves more than one feature. Because such a long chunk is probably more difficult to encode than two smaller two-object chunks, we separated these long chains by pairing the objects, as follows: □■ and ■□. This method was less strict for scoring participants' performance. For instance, based on the □■■□ study sequence, a □■ chunk would be identified in □■■ (the participant's response sequence), whereas no chunk would be credited to the participant if □■■□ was coded as the only available chunk in the study sequence. In comparison to the first method for identifying the chunks in the sequences, the number of objects recalled in the similarity condition increased from 1.13 to 2.07 using strict-position scoring ($t(66) = 24, p < .001, \eta_p^2 = .90$), and from 1.34 to 2.5 using free-position scoring ($t(66) = 31.1, p < .001, \eta_p^2 = .94$). This method could not, however, affect the rule-based condition much, because, by construction, none of the chunks could include more than two objects in this condition.

Overall, the number of objects recalled was greater in the rule-based condition, meaning that the participants were more inclined to encode a greater number of smaller chunks than a smaller number of chunks made of a longer chain of similar objects (even though the long chain could

be split into several contiguous pairs).

To come back to the analysis of the transitional error probabilities (now that we know that the analysis does no longer favors the identification of chunks that are rule-based), we again divided the frequency of correct transitions by the total number of times that the first item was correctly recalled for a given pair of adjacent items. We then averaged the frequencies of the within-chunk and between-chunk transitions across participants in order to run a repeated-measures ANOVA using order (rule-based *vs.* similarity-based) and type of pair (within-chunk *vs.* between-chunk) as within-subject factors. This analysis showed that a greater number of transitions were correctly made (in comparison to Table 1) for the within-chunk transitions ($p = .64$) than for the between-chunk transitions ($p = .48$), $F(1, 66) = 245, p < .001, \eta_p^2 = .80$, and that more transitions were correctly made in the rule-based condition ($p = .60$) than in the similarity-based condition ($p = .52$), $F(1, 66) = 31, p < .001, \eta_p^2 = .32$, with no interaction between the two factors. Overall, there were 2178 correct transitions versus 1260 transitional errors in the within-chunk condition, but only 1432 correct transitions versus 1685 transitional errors in the between-chunk condition, corresponding to an odds ratio of 1.4.

Discussion

We explored the ability of untrained participants to increase their immediate memory by parsing sequences of objects into newly-formed chunks. The main reason for exploring chunk formation in immediate memory is that a chunk is too often thought of solely as a product of an already-formed long-term representation. Showing that recoding can occur very rapidly in immediate memory is a different undertaking. This idea may seem merely intuitive, but no other model can give a precise quantitative account of how information is apprehended to determine exactly what patterns can be extracted rapidly to increase memory capacity. Chunking demands a

more principled explanation than one simply saying that individuals demonstrate chunking when they group items together while performing a memory task. The chunking memory-span task presented in this study provides a way to explore the formation of a chunk in immediate memory, without relying on chunking that has become familiar via repeated exposure.

Our theory was that a compressibility metric could estimate the opportunities available to participants for immediate recoding of the to-be-recalled material. A first result showed a capacity of about three objects for the most noncompressible sets of objects, at a 50% threshold (see Fig. 4, which corresponds to the lower limit of the 4 ± 1 estimate; Cowan, 2001). This result is consistent with the literature (e.g., Broadbent, 1975 ; Chen & Cowan, 2009), despite Cowan's slightly higher estimate (2001). However, in line with our prediction, this capacity limitation is overcome whenever a form of relational information can be computed, that is, when compressible sets of objects are used. The proportion of sequences retained increased with compressibility, particularly when presentation order facilitated the recoding of the sequences (up to about four objects at a 50% threshold; see Fig. 4 for the rule-based order). Overall, this finding showed a minimal span of about three objects that increased monotonically as a function of sequence compressibility.

More refined analyses showed that the objects were most often chunked into pairs, in an incremental fashion. The results also showed that recoding occurred more often in the rule-based order (which favors short, well-clustered logical chunks of two objects) than in the similarity-based order (which favors longer chunks of very similar objects). The dissimilarity condition was less likely to induce chunking because it tended to cluster the objects individually. However, it is more difficult to tell exactly what a chunk or rule is in immediate memory. More specifically, because our method targets a process that occurs before any consolidation process, there is no clear-cut empirical measure of the presence of chunks or rules. Beyond that, however,

by calculating the transition-error probabilities and analyzing how chunks were incrementally encoded, we do provide clues as to the nature of the chunks formed.

Chunking in the present study probably emerged in immediate memory thanks to the identification of elementary structures of repetition in the input stream. This is why the present study goes beyond those that have used digit sequences to study chunking (Bor & Owen, 2007 ; Mathy & Feldman, 2012), insofar as regular digit sequences such as 123 may already have been represented in the participants' long-term memory (see Jones & Macken, 2015). Also, the chunking memory referred to in our tasks does not correspond exactly to Potter's (1993, 2012) notion of conceptual short-term memory. In her theory, short-term memory operates by allowing meaningful patterns stored in long-term memory to be unconsciously identified in short-term memory. Our results are closer to those of a previous study in which untrained participants could improve memorization of a set of spatially-grouped objects organized into smaller units by using hierarchically nested levels (Feigenson & Halberda, 2008).

However, we have not lost sight of the fact that some individual strategies devoted to describing the sequences can escape predictions. For instance, we took the risk that some verbal processes usually shown to be critical to short-term memorization (Burgess & Hitch, 1999 ; Baddeley et al., 1975 ; Estes, 1973 ; Zhang & Simon, 1985) are left uncontrolled in our study because the items can be recoded verbally. Still, we intentionally let the participants recode (verbally or otherwise) our visual-based material because we thought that implementation of articulatory suppression could also affect attention and subsequent encoding of regularities. It was our goal to let the participants encode the objects freely.

The next sections evaluate competing theoretical accounts of the compression effects we found. We will discuss in detail the under estimated role of immediate memory in the formation

of chunks and why the compressibility of information is a good theory to capture grouping effects. We begin with the idea that although interference-based models can easily account for forgetting, they cannot account for the potential chunking processes at play in our data. Indeed, the non-distinctiveness of the stimuli can paradoxically be both detrimental to retaining similar items (when it is considered to produce an interference-based forgetting process) and beneficial to grouping (in our study, when it produces a chunking process).

Interference-based models and local distinctiveness account

We believe that short-term memory models based on the stimulus distinctiveness do not explain our presentation-order effects. These models contend that similar items interfere with each other (Lewandowsky, Duncan, & Brown, 2004 ; Nairne, 1990, 2002 ; Oberauer & Kliegl, 2006). At best, these models predict no difference between presentation orders, because the overall similarity of items is identical in our three orders. At worst, these models incorrectly predict poorer performance in the similar condition than in the dissimilar condition because of the possible negative effect of similarity between adjacent items.

In contrast to both of these predictions, we correctly predicted better performance in the similarity-based order than in the dissimilarity-based order because the particular kind of similarity we presented allowed for easier compression of the material based on feature continuity between adjacent items. This result supports a more general finding that similarity can be either beneficial or detrimental to memory (Gupta, Lipinski, & Aktunc, 2005 ; Hunt & McDaniel, 1993 ; J. S. Johnson, Spencer, Luck, & Schöner, 2009 ; Lin & Luck, 2009 ; Nairne & Kelley, 1999). Moreover, we correctly predicted the best performance when the materials lent themselves directly to explicit rules of transition. The similarity condition did have longer potential chains of similarity between contiguous items, but the basis of the similarity kept shifting, so

participants were generally unable to use those contiguities as consistently as they could in the rule-based condition. The logical content of our task might explain why an associative process based on similarity occurred less often in our data than a rule-based process did (Sloman, 1996).

For instance, the SIMPLE model (Brown, Neath, & Chater, 2007) suggests that the retrieval of items in memory is a discrimination process at all periods of time (short-term and long-term). In SIMPLE, items may be hard to retrieve when their temporal distance is short, for the same reason that similar items are hard to identify. SIMPLE usually captures the idea that items are more discriminable when they occupy isolated locations on the temporal dimension. Here in our study, it was adapted to capture the idea that items are more discriminable when they are both physically and temporally isolated. The following equation derived from context models (Nosofsky, 1984, 1986) is simple enough to discriminate the effect of our three presentation orders:

$$d_{ij} = \sum_{a=1}^{n} |x_{ia} - x_{ja}| \tag{1}$$

The preceding equation can be used to evaluate both the physical distance and the temporal distance between two items in a pair. For instance, let $d_{ij} = 2$ when a ■ is presented one second after a □ (in this case, the total distance is: one feature + one second = 2 differences). The following exponential decay function can be used to relate stimulus similarity to psychological distance (Nosofsky, 1986 ; Shepard, 1987):

$$\eta_{ij} = e^{-d_{ij}} \tag{2}$$

Finally, SIMPLE assumes that the discriminability D_i of a memory trace is inversely propor-

tional to its summed similarity to the n memory traces:

$$D_i = \frac{1}{\sum_j^n \eta_{ij}} \tag{3}$$

The probability of recalling item i in its correct temporal position is simply its discriminability.

It appears that our results cannot be modelled by SIMPLE, in which performance may be accounted for by an interference process that would favor the dissimilarity-based order. Looking at Fig. 3, our simulation showed that the overall discriminability ($\sum D_i$) of the six items in the rule-based order was 3.90 (.69, .64, .62, .62, .64, .69 for the respective six items respectively, summing to 3.9), versus 3.27 for the similarity-based order and 4.79 for the dissimilarity-based order, when only the items' physical dimension was taken into account. Adding the temporal dimension did not change this ranking and we simply obtained larger overall discriminability values: 5.09, 4.80, and 5.50, respectively. In this interference-based model, the dissimilarity-based order is predicted to produce maximal local distinctiveness (hence better recall), whereas the similarity-based order is predicted to produce maximal interference in memory and subsequently a negative effect on memorization. The rule-based model has an intermediate status in the model, although our results clearly showed its superiority. SIMPLE cannot account for how chunking operates in immediate memory with the present data (even if the predictions are reversed to consider similarity as beneficial, the rule-based order still has an intermediate status).

Compressibility of information

In the present study, recall performance was significantly enhanced for compressible sequences as compared to less compressible ones. Brady et al. (2009) introduced statistical regularities

in some sequences of visually presented stimuli in order to study the effect of redundancy on memorization. The authors reported that the observers recalled more information when the stimuli exhibited regularities. They explained this advantage by the fact that several items associated in memory formed a single chunk, leaving more room in memory for storing other stimuli. A compression process using statistical regularities (Brady et al., 2009) can be perfectly modeled with a minimum description length (MDL) type of approach (Rissanen, 1978). MDL is based on an algorithmic notion of compressibility that seems more versatile than Feldman's (2000). However, the MDL approach seems to offer no advantage for recoding a single short sequence such as the ones used in the current study. A simple example is the difficulty applying MDL to a short sequence such as 12349876. In this case, its recoding into $a = 1234$ and $b = 9876$ does not lead to any compression (i.e., the length of the representation "$a = 1234$ and $b = 9876$, ab" is greater than the original sequence, because the redescription process needs to be taken into account), whereas it can be clearly useful for a sequence that is slightly longer, such as 123123123123123123, for which we can use a shorter representation $a = 123, aaaaaaa$ (the $a = 123$ part corresponds to what is written in the lookup table, and $aaaaaaa$ corresponds to the recoded sequence, but combined, it is shorter than the original string). The same advantage for $a = 123$ occurs if 123 is repeated many times in the long-term mixed in with other sequences (e.g., $123, 4, 123, 123, 5, 6, 2, 1, 123, 123, 123, etc.$). We concentrate here on a practical version of MDL. However, MDL is ideally an algorithmic approach to compression that does not have to rely on frequently repeated patterns, and as such, it is not impossible that MDL could recode a small individual sequence like 12349876 based on a particular appropriate programming language.

We agree with Brady, Konkle, and Alvarez's (2009) suggestion that chunking can be used as an approximation of an algorithm of psychological compression. However, our results go beyond

that earlier work by showing that a short amount of time is sufficient to allow a compression process to occur on the fly. Our interpretation is that compression can be efficient if it makes use of primitive repetition structures, which can be detected and used rapidly in immediate memory. As such, Feldman's model of compressibility is more suited than MDL for capturing the recoding of single short sequences like the ones in the current study. For instance, for the set ■□◯●, compressibility is very limited and requires a more extensive description in working memory than a more compressible set. When almost no compression is possible, the representation must contain almost as many features as in the original list of objects.

A last point concerns the relationship between compressibility and presentation order. When the members of a compressible set (e.g., one where FC = 1, such as △▲△▲) are unordered, their memorization can be facilitated by using the rule "triangles". However, the memorization of the ordered objects △▲△▲ requires a longer description than "triangles". Therefore, a compressible category that is not ordered in a regular way can be as difficult to memorize as a noncompressible one. This is noticeable in Fig. 4 in which the mean data point for FC = 10 (this data point is represented by the black curve where length = 4) is very close to the mean data point for the green dissimilarity curve. However, when no compressibility is allowed in the first place, order does not matter anymore.

Conclusion

One common idea is that working memory simply uses pointers to retrieve chunks in long-term memory, and that chunks are only consolidated pieces of information stored in long-term memory. We believe that we observed a chunking process that seems to result from the temporary creation of new representations. Recall performance in the current study was as small as about three items, but the total increased when the sequential order of the items was favorable

to the formation of new chunks, i.e., compression of the sequence based on rules discovered within it. One potential strong conclusion is that these variations in performance mean that chunking processes can operate during short-term memorization. It follows from this conclusion that chunks must not be considered solely as permanent groupings of information in long-term memory. It is possible that our participants spontaneously compressed information without much mediation of long-term memory chunks, since it is probable that none of our participants had ever encountered the particular sequences of objects seen in our experiment.

The experiment presented in this paper allows us to draw the conclusion that information can be compressed in a few seconds in immediate memory by capitalizing on the regularities within the materials presented. Our research challenges us to consider that the notion of compression in immediate memory is central to defining grouping, chunking, and probably learning.

Appendix

Notes on Feldman's (2000) compressibility metric

To use Feldman's (2000) compressibility metric, the stimuli must be split into categories. Categories (or simply groups of stimuli) from 1 to 8 stimuli were drawn from the initial set of 8 stimuli. The k chosen stimuli represented the class of to-be-recalled stimuli, while the $8 - k$ remaining objects represented the opposite class that did not have to be recalled after stimulus presentation but were shown in the recall screen. Using all of the eight stimuli in the procedure was thought to preserve the complexity metric on which the present study was based. Had the participants been only requested to reconstruct the order of the chosen stimuli (e.g., □ followed by ■) without contrasting them to the opposite category, the participant would only have to encode the color feature to succeed (e.g., using a "white first" representation of the sequence). For instance, to follow the metric that was first thought to account for categorization processes, the □ ■ sequence, needs to be contrasted to the opposite class (i.e, △ ▲ □ ■ △ ▲) by the participants. In this respect, the minimal representation of the category is "small squares", and the right order ("white first") applies specifically to the "small squares" category. The minimal representation ("small squares") that inter-relates the individual stimuli is the compressibility of the category because it is shorter than "small white square or small black square". The minimal representation that includes order information is "small squares, white first". Note that it is principally because participants need to remember a sequence in the correct order that the task goes beyond what a categorization task would require.

Figure 6 gives Feldman's (2000) measure of complexity (FC) for each category, which was used as a measure of compressibility. This measure indicates the number of features in the shortest logical formula allowing one to describe the stimuli, which are indicated by a black dot. The

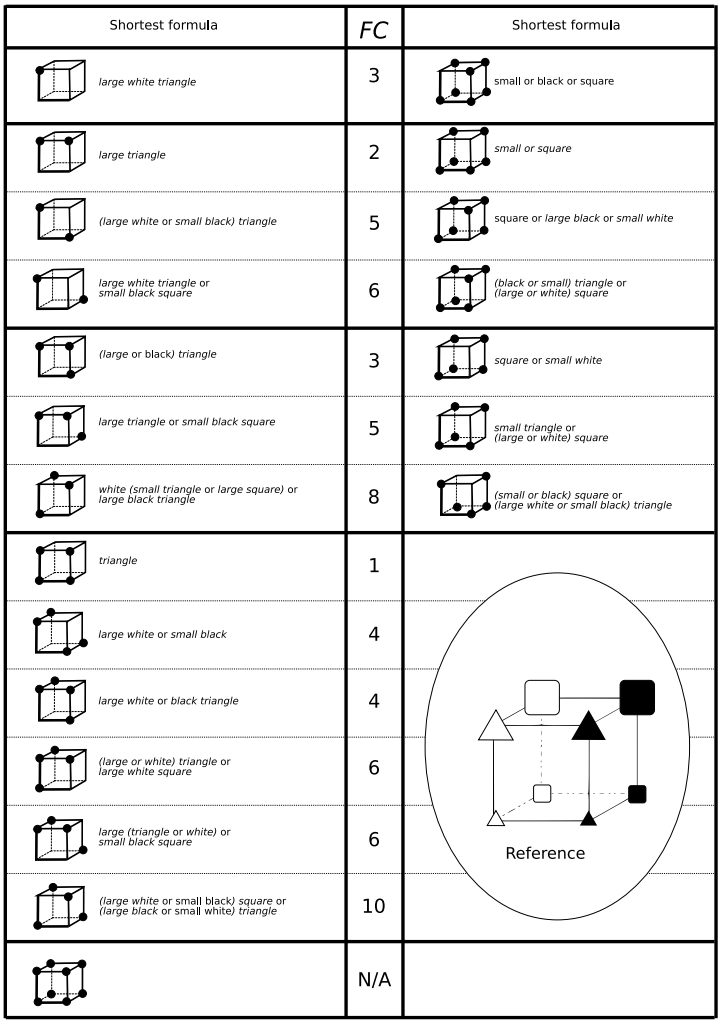


Figure 6 – Feldman's (2000) minimization of Boolean complexity. *Note*. Column FC indicates Feldman's (2000) measure of logical compressibility for each of the categories in the left or right column.

black dots positioned on the cubes illustrate the relationship between the objects that could be used by the participants to associate the stimuli to be recalled in our experiment. Each of the stimuli selected for a category were presented sequentially and their location remained constant in the middle of the screen, so no reference was made during the experiment to the spatial organization shown in this figure (this was done here to bring out the similarity structure between objects for the reader). For instance, the eighth cell on the left indicates that four triangles were used to construct the memory list (see bottom right cube above 'Reference' for retrieving the stimuli to which the black dots in the eighth cell on the left are pointing). The "triangle" category representation would be sufficient for free recall of the list.

FC is indicated in the middle column because the FC numbers for a given line fit both the category structures on the left and those on the right. Note that FC is not confounded with sequence length for sequences equal to 2, 3, 4, 5, or 6 objects. However, for the categories made of 1, 7, or 8 objects, no variations in complexity could be manipulated. When a set was composed of eight objects, categorization was no longer involved. In this particular case, the shortest formula was trivial ("all objects") and it was associated with a 'Not Applicable' (N/A) FC value.

Troisième partie

Les tâches d'empan simples et complexes ne présentent pas de différence fondamentale

There is no fundamental difference between simple and complex span tasks

Fabien Mathy

Université Nice Sophia Antipolis

Mustapha Chekaf

Université de Franche-Comté

Nelson Cowan

University of Missouri-Columbia

Abstract

Researchers have suggested that simple span tasks and complex span tasks measure different theoretical constructs. Simple span tasks are tasks commonly used to measure short-term memory, while complex span tasks are usually considered typical measures of working memory. Because complex span tasks were designed to create a demanding concurrent task, the average span is usually lower (4 ± 1 items) than in simple span tasks (7 ± 2 items). One possible reason for the higher span of simple span tasks is that participants can take profit of the spare time between the stimuli to detect and recode regularities in the stimulus series (in the absence of a concurrent task), and such regularities can be used to pack a few stimuli into 4 ± 1 groups. We predicted that the respective spans of these two types of tasks may be equal (at around 4 ± 1) when regularities are absent from the stimulus lists. Using a simple span task with visual stimuli offering the smallest amount of relational information so as to make the stimulus lists incompressible and thus, non-chunkable, we showed with a large sample of participants ($N > 300$) that the average simple span was less than three objects. This result confirmed that simple spans can be diminished to a large extent by the use of incompressible materials which were hypothesized to block a chunking process. However, observation of the spans in the non-chunkable vs. chunkable series refuted the idea that compressibility is important solely in simple spans. Indeed, the chunkability of the material still operated in the complex span tasks, meaning that chunking can occur even in complex span tasks.

Working memory (WM) capacity is often assessed by complex span tasks (e.g., reading span, operation span; see Cowan et al., 2005). Complex span tasks interleave a memory task (requiring one to remember a set of items in the correct order) with a secondary processing task (e.g., judge the correctness of equations). The dual nature of complex span tasks is central to Baddeley's (Baddeley & Hitch, 1974 ; Baddeley, 2000) model, which distinguishes a storage component and a central attentional control system, but also to more recent models (Barrouillet et al., 2004 ; Engle, 2002). One advantage for conducting complex span tasks is that they are predictive of intelligence (A. R. Conway et al., 2002 ; Cowan et al., 2005 ; Unsworth & Engle, 2007b), which notably requires both memorizing and manipulating several pieces of information to solve a problem.

However, a great many daily activities that require working memory often focus on a single task, such as remembering a telephone number or the names or faces of people one has just met. Traditionally, simple memorization of a list of items (such as numbers) has been assessed by simple span tasks. Such tasks have produced very stable estimates since the 1920's (Gignac, 2015), i.e., around the well-known 7 ± 2 capacity limit (Miller, 1956), despite the fact that alphanumerical materials rely on the linguistic experience of the rememberer (Jones & Macken, 2015). Simple span tasks are less demanding than complex span tasks because no secondary task is used, and for practical reasons they are very often used in intelligence tests such as the Wechsler's, which are paper-pencil tests. Put simply, simple span tasks are commonly considered typical tasks for measuring short-term memory (which is considered passive and dedicated to item memorization), while complex span tasks are usually considered typical measures of working memory (which is assumed to involve information processing).

Because complex span tasks were designed to include a concurrent task, the average span is

usually lower than in simple span tasks, that is, around 4 ± 1 items (Cowan, 2001). As a result, but maybe erroneously (Unsworth & Engle, 2007b), research has suggested that simple span tasks and complex span tasks measure different theoretical constructs (Aben et al., 2012) or different processes (A. R. Conway et al., 2002 ; Unsworth & Engle, 2006).

One possible explanation for the higher span of simple span tasks is that participants can detect regularities in the stimulus set, which are used to pack a few stimuli into groups (Chekaf, Cowan, & Mathy, 2016 ; Mathy & Feldman, 2012 ; Miller, 1956). One example of regularity in a digit series would be 1-9, which tends to point to a known date. Two letters can also easily present a regularity if they evoke a part of a word, as well as several visual stimuli that share a feature based on localization (a cluster for instance). This would imply that some information is processed in simple span tasks as well as in complex span tasks (except that the processing would be more directly devoted to storage in simple span tasks, instead of being directed away from storage as in complex span tasks). Forming three or four groups is in fact sufficient to artificially increase capacity to 6 or 7 items (e.g., 198452 = '19'-'84'-5-2, or '1984'-5-2), but there is greater chance that this process occurs when there is no concurrent task. To manipulate the compression process in our task, we used basic visual stimuli (such as colored shapes), such as ■ ■ ▲ ▲. This series is regular in that the color dimension can be used to recode the objects and because the shapes can be grouped by pairs. The idea developed in the current study is that we should observe a similar capacity (i.e., 4 ± 1 items) for simple and complex span tasks if stimulus processing is hindered in both cases. Such an idea has already been tested using very rapid presentations, as in the change detection paradigm (Luck & Vogel, 1997), for which a similar storage capacity limit of 3 or 4 is generally observed. Our goal was to design a simple span task using materials complicated enough to hinder rapid re-encoding of information. We derived our task from the categorization domain, which has shown that information can hardly been re-

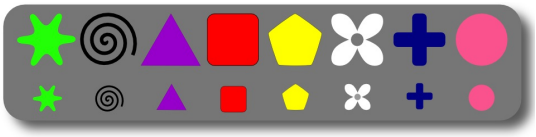
encoded in circumstances where information is not compressible (Feldman, 2000, 2003 ; Lafond et al., 2007 ; Vigo, 2006 ; Goodwin & Johnson-Laird, 2013 ; Bradmetz & Mathy, 2008). We think that this domain which is generally connected to learning can be helpful to study how memory functions with complex visual material. One difference between categorization experiments and the present experiment is that our stimuli were presented only once for immediate recall. The current study differs from our previous study (Chekaf et al., 2016) in that it manipulates the type of span task to determine whether compressibility occurs only in simple span tasks or also in complex span tasks, whereas the previous study mostly focused on the linear effect of compressibility on recall in simple span tasks. This previous study showed that the amount of information retained in short-term memory was maximal for the most compressible sequences, but the procedure only involved simple span tasks.

In the following, we use "compressibility" and "chunkability" almost seemingly as synonyms, but the relationship is thought to be actually causal. Previous work has shown evidence that memoranda tend to be chunks when more compressible. The idea is that individuals can better optimize their limited capacity in working memory when there is opportunity to reduce the sum of information by re-encoding stimuli into shorter representations (Chekaf et al., 2016 ; Mathy & Feldman, 2012). Our first experiment directly compared a simple span task to a complex span task, using compressible *vs.* incompressible sequences of stimuli (nonverbal objects of different shapes, sizes, and colors). We expected to find that the compressibility of the materials would make a difference only for simple spans, under the assumption that compression processing was not possible in complex span. Two aspects of the results were surprising. First, we found that compressibility was quite important for both simple and complex span tasks, suggesting that the processing task in complex span does not prevent compression processes. Second, the level of simple span with incompressible materials was even lower than we expected, so a second

experiment was carried out to check that result in simple span. We discuss the observation that compression is more ubiquitous than predicted, and propose that unlike rehearsal (Cowan et al., 2005), chunking can presumably occur even in complex span tasks.

Experiment 1

To prevent the compression process in our new simple span task, we limited the relational information in a set of basic visual stimuli (such as colored shapes), using simple colors and simple shapes. More specifically, we used lists of nonverbal, categorizable, artificial three-dimensional stimuli, with two-valued/Boolean discrete features for the shapes, sizes, and colors shown in Figure. 7. For a given three-dimensional set, we selected sets with stimuli of the lowest relational information, a manipulation that was hypothesized to prevent chunking. Let us illustrate with the three-dimensional set of objects: □ ■ □ ■ △ ▲ △ ▲. A compressible subset of four objects would be: ■ ■ ▲ ▲, because the color dimension is sufficiently diagnostic to discriminate black objects from white objects. A stimulus list based on this subset offers the possibility of re-encoding the sequence using the simple rule 'black'. The sequence ■ ■ ▲▲, in which order matters, can be described by a simple rule using the 'black' feature (to recode the entire subset, regardless of order) and the 'square-triangle' order which can be combined with a 'large-first' description within each shape. By contrast, a less compressible subset of objects would be: □ ■ ▲ △. The heterogeneity of these four objects that makes the category structure complex can be measured by the difficulty of compressing information, accounting for the difficulty of recoding the stimuli into a more compact representation (Bradmetz & Mathy, 2008 ; Mathy & Bradmetz, 2004 ; Feldman, 2000, 2003 ; Goodwin & Johnson-Laird, 2013 ; Kruschke, 1992 ; Lafond et al., 2007 ; Mathy, 2010 ; Sloutsky, 2010 ; Vigo, 2006). In other words, there is no simple, hierarchical rule that explains the sequence of shapes/colors of this



Trial	Length	Inter-item distance	Category strucure	To-be-remembered list
Non-chunkable category structure				
1	1	0	□	
2	1	0	□	
...				
5	2	3	□	
6	2	2	□	
...				
9	3	5	□	
10	3	4	□	
...				
13	4	7	□	
14	4	6	□	
Chunkable category structure				
1	1	0	□	
2	1	0	□	
...				
5	2	1	□	
6	2	1	□	
...				
9	3	3	□	
10	3	3	□	
...				
13	4	4	□	
14	4	4	□	

Figure 7 – (Top) A sample of stimuli based on eight shapes, eight colors and two types of sizes. (Bottom) Table showing a sample of trials in Exp. 1 and in Exp. 2. The first half of the table shows the sequences of the non-chunkable condition used in both Exp. 1 and in Exp. 2. The second half of the table shows the sequences of the chunkable condition used in Exp. 1 exclusively. For each sequence length, we chose to represent only two trials. Dimension values were chosen randomly for each trial. For instance, the given category structure □ ■ □ ■ (i.e., square, small first, white-black within sizes) could be converted into either △ □ △ □ (i.e., white, small first, triangle-square within sizes), or ■ ▲ □ △ (i.e., large first, square-triangle within colors), etc. The preceding example only involve the dimension values square, triangle white and black, but again, the dimension values were actually randomly picked among the 8 shapes and the 8 colors shown in the top panel.

subset. More homogeneous category sets produce a lower information load and as such, they are more compressible and can be easily re-encoded (or "chunked") to facilitate recall (Chekaf et al., 2016).

Hereafter, we call the simple compressible homogeneous sequences "chunkable" and the complex sequences "non-chunkable" (or "less-chunkable" when more convenient). Accordingly, four conditions were constructed: *SimCh*, a simple span task using chunkable material; *ComCh*, a complex span task using chunkable material; *SimNon*, a simple span task using non-chunkable material; and *ComNon*, a complex span task using non-chunkable material.

We predicted an interaction between task and chunkability. The predicted pattern of results was $SimCh > ComCh$ (meaning that performance would be greater in the simple span task using chunkable material than in the a complex span task using chunkable material) but $SimNon = ComNon$ (spans would be equal using non-chunkable material in simple and complex span tasks). We predicted that the simple span task could only have a beneficial effect on recall when some of the information could be re-encoded (i.e., $SimCh > ComCh$), while such a benefit could not occur when no information (or little information) could be re-encoded (i.e., $SimNon = ComNon$).

Another stronger, direct prediction was $ComCh = SimNon$, with the idea that a complex span task offers no opportunity to recode the regular patterns in the chunkable condition because attention is directed away during the concurrent task, while participants cannot profit from spare time in a simple span task to recode information when there is too little information to be recoded in the non-chunkable condition.

A more modest prediction was that there would be an interaction, with a greater difference between the simple span task and the complex span task for chunkable patterns than for non-

chunkable ones, that is, $SimCh >>> ComCh$ but $SimNon > ComNon$.

One last goal was to test the size of the interaction (if present) to compare the quantity of material chunked in the four conditions. We will run a multiplicative analysis to test whether the gain between the non-chunkable and the chunkable conditions is similar in the simple and complex span tasks. The reason would be that the incompressible condition still contains little information that can be re-encoded. Finding a multiplicative scale would mean that chunking can also occur in complex span tasks. For instance, if twice as many objects are encoded when the material is chunkable as when it is not, regardless of the type of span task, we could observe an increase of four to eight times with chunkability on the simple span task, and two to four times with chunkability on the complex span task. These numbers would correspond to a constant ratio of $8/4 = 4/2 = 2$ more objects encoded when material is compressible in both tasks. A constant ratio across tasks would tend to indicate that the ability to recode information is constant across tasks (although more difficult in general in complex span tasks) in spite of the interaction. A smaller ratio for the complex task would indicate not only a strong interaction but also an ability to recode information far greater in simple span tasks.

Participants

Ninety-four students ($M = 23$ years old, $sd = 5.3$) enrolled at the Université Nice Sophia Antipolis volunteered to participate in the experiment.

Stimuli

Our stimuli varied according to three two-valued/Boolean dimensions (shape, size and color, the three dimensions typically used by category learning researchers to build canonical stimulus sets; Love & Markman, 2003). We used only two values per dimension within each trial (Figure 7).

For each trial, a random combination of two shapes (among eight different ones), two colors (among eight different ones), and two sizes made a set of eight possible objects. We restricted the size dimension to two different values (large *vs.* small, i.e., 280 × 280 pixels *vs.* 140 × 140 pixels) across lists because participants had trouble identifying intermediate values during our pre-tests. The use of eight shapes, eight colors, and two sizes was sufficient to generate 1568 possible sets of eight objects, which limited proactive interference between trials.

The participant did not know in advance which of the dimensions would be the most relevant to the categorization process. Dimension values were chosen randomly for each of the lists presented, so as to vary the possible combinations of dimensions (shapes, sizes, and colors) across lists, while preserving the same category structure (shown in Figure 7). The probability that a participant would come across two identical sets of features between two lists during the experiment was assumed to be very low.

Procedure

The experiment was a 2 × 2 within-subject design. Each participant attempted all four blocks (chunkable simple span task, non-chunkable simple span task, chunkable complex span task, non-chunkable complex span tasks), the order of which was counterbalanced across participants (i.e., 24 possible orders; 96 participants were needed to perfectly balance the design). Each block comprised several lists of stimuli and recall occurred after each list. The participants were informed that they were required to memorize, in correct order, each list of stimuli. A list of stimuli (e.g., a small blue square and a large blue square) was chosen from a random combination of two shapes (e.g., all the stimuli resulting from the combination of small *vs.* large, blue *vs.* red, and square *vs.* circle objects). The stimuli in a given sequence were displayed serially in the center of the screen for one second each (e.g., for a two-stimulus list, a small blue

square followed by a large blue square). Difficulty of each sequence was estimated following the compressibility metric described by Chekaf et al. (2016). After the list of items was presented, the response screen showed the whole set of eight objects from which the subset had been selected. The response screen showed in randomly-determined positions eight response choices: the k to-be-recalled stimuli and the $8 - k$ remaining distractor objects. For Trial 14 of the non-chunkable condition in Figure 7, the recall screen would have also included a large green triangle, a small purple triangle, a small green circle, and a large purple circle as the new items. Trial 14 of the chunkable condition would include the same objects but in red. The chunkability of the memoranda was therefore voluntarily correlated with retrieval demands of the trials. Following the example, the new items of the non-chunkable trial 14 are more like lures than the respective chunkable trial because the features overlap with those memorized. Only this manipulation respects the compressibility metric, because without a response screen using more objects than the memoranda, there would be no need, for instance, to encode the feature 'blue' of Trial 14 of the chunkable condition. The participant made selections by clicking on the objects to recall the items in the correct order. This recall procedure is similar to that of the visual short-term-memory serial report task (Avons & Mason, 1999 ; Smyth et al., 2005). The stimuli were underlined using a white bar when the user clicked on them. There was no timing constraint for recall. The participant could move on to the next sequence by pressing on the space bar.

The lists were displayed using ascending presentation of length (length varied progressively from 1 to 8 items), as in the digit spans used in neuropsychological tests. Trial length 1 was only used as a warmup. For instance, our experiment used the same number of repetitions per length as the digit span of the WISC or WAIS. A block automatically stopped after four errors within a given list length. Participants were given four trials per length L. They were also

informed that the first three trials in each block would be treated as practice trials and then discarded from the analysis. After this warmup, there was four trials per list length in each condition.

When the task was a complex span task, we used the operation span (OS) task procedure. In OS, participants are required to perform mathematical operations between memory items (see Conway et al., 2005; Kane et al., 2004). An equation was displayed on the screen (e.g., "7 + 2 = 10") before each to-be-remembered item was presented. The participant had three seconds to judge the equation by clicking a button (true or false), before the next item was displayed. This concurrent task was thought to prevent participants from chunking freely.

For the non-chunkable simple span, for a given list length, the most incompressible lists alternated with less incompressible lists; otherwise, the inter-item distance would have exhibited too much regularity across the experiment (i.e., a ubiquitous three-feature inter-item distance). For instance, in Figure 7, Trial 10 of the non-chunkable category-structure panel shows the most incompressible three-object set, with a first 2-feature difference (size and color, between the little white square and the large black square) followed by a second 2-feature difference (size and shape, between the large black square and the small black triangle), whereas Trial 9 shows a more compressible 3-object set, but ordered using a 3-feature difference followed by a 2-feature difference to make the grouping process harder. The inter-item distance is convenient to describe the relationships between features, but Feldman (2000, 2003) describes more precisely how the features can be redescribed to compress the sum of information in each set of objects (the compression process is not always related to inter-item distance). For instance, "small white square, small black square" can be reduced to two features ("small square"), whereas "small white square, large black triangle" cannot be reduced to less than six features. This measure of compressibility only serves here to predict the chunkability of a category set. Overall, all of the

category structures of a given length were chosen to be less compressible in the non-chunkable condition than in the chunkable condition.

Scoring

A value of .25 was scored for each perfectly correct serial report of all the memory items (all-or-nothing method). A participant recalling only 3 out of 4 sequences of one object would be granted a span of .75. When a subject obtained 4, 3, and 2 trials correct at lengths 1, 2 and 3 respectively, then the span was equal to $9/4 = 2.25$.

Results

Effect of task procedure and category-set complexity

An ANOVA was performed with task procedure (simple *vs.* complex span task) and category-set complexity (chunkable *vs.* non-chunkable) as repeated factors, using the mean span as the dependent variable. This analysis yielded a significant effect of task procedure ($F(1, 93) = 421.85, p < .001, \eta_p^2 = .82$), a significant effect of category-set complexity ($F(1, 93) = 268.22, p < .001, \eta_p^2 = .74$), and a significant interaction between the two ($F(1, 93) = 40.29, p < .001, \eta_p^2 = .30$) due to *SimCh* >>> *ComCh* and *SimNon* > *ComNon*. In the simple-span and complex-span conditions, the mean spans were 3.45 ($sd = .77$) and 1.94 ($sd = .70$), respectively. Overall, the mean spans when the lists of objects were chunkable and non-chunkable were 3.35 ($sd = .83$) and 2.08 ($sd = .49$), respectively. The spans in all four conditions are shown in Table 2 and Figure 8.

Analyses of the simple effects showed that *SimCh* > *ComCh* ($t(93) = 16.06, p < .01$) and *SimNon* > *ComNon* ($t(93) = 14.88, p < .01$), but also that *SimCh* > *SimNon* ($t(93) =$

$15.23, p < .01$) and *ComCh* > *ComNon* ($t(93) = 8.55, p < .01$). This partly disconfirmed our main prediction that *SimCh* > *ComCh* and *SimNon* = *ComNon*. Further analysis also refuted the stronger prediction that *ComCh* = *SimNon* ($t(93) = 2.97, p < .01$). To test the last prediction that *SimCh* >>> *ComCh*, in comparison to *SimNon* > *ComNon*, we compared the difference *SimCh* − *ComCh* to the difference *SimNon* − *ComNon*. The difference between the two was significant ($t(93) = 6.35, p < .01$).

Overall, one surprising result is that the average span for condition *SimNon* was less than three objects ($t(94) = 8.4, p < .001$), which is lower than the lower bound of the 4 ± 1 capacity limit.

Multiplicative chunking effect test

Despite the larger difference between *SimCh* and *ComCh* than between *SimNon* and *ComNon*, we can conclude that chunkability had roughly the same effect in the simple and complex-span task conditions. We defined a ratio equal to the chunkable span divided by the non-chunkable span for each type of task, and we obtained the following values: Simple span: 4.05 (chunkable)/2.49 (non-chunkable) = 1.63; complex span: 2.2 (chunkable)/1.45 (non-chunkable) = 1.52. Chunkability in the lists of objects had a multiplicative effect on recall, insofar as participants recalled about 1.5 times more items when there were regularities than when there were not.

Chunking scores were calculated for every participant and submitted to a Bayesian analysis. Using JASP [8], we tested the null hypothesis that the span ratios of the simple and complex tasks would be the same. The appropriate alternative hypothesis was that the ratio for complex span tasks is lower than the ratio for simple span tasks, i.e., that stimuli in simple span tasks are

8. JASP. (2014). Retrieved from http://jasp-stats.org/

Table 2 – *Mean span (and standard errors) in Experiment 1, by procedure (simple vs. complex span tasks) and category set complexity (chunkable vs. non-chunkable).*

	Chunkable	Non-Chunkable
Simple Span	*SimCh*	*SimNon*
	4.05(.11)	2.49(.06)
Complex Span	*ComCh*	*ComNon*
	2.2(.09)	1.45(.06)

more highly chunkable. The Cauchy Prior with a width of .707 was used to test the alternative hypothesis (the program's default). With that alternative hypothesis (using just half of the Cauchy Prior), the Bayes factor was 12.69 in favor of the null hypothesis over the alternative hypothesis. A less theoretically-guided test would be a two-tailed test, the alternative now being that the ratios could differ in either direction. With that alternative hypothesis, the data came out with a Bayes factor of 7.58 in favor of the null hypothesis. Both exceed the ratio of three that seems to be a standard convention for a sufficiently decisive finding. This other surprising result leads us to believe that chunking in immediate memory is a deep process that is resistent to concurrent tasks, at least deeper than what our initial hypotheses suggested [9].

Experiment 2A

The goal of Experiment 2A was to be able to generalize our surprising finding that the average span for the simple span task and incompressible material (i.e., 2.49) was found to be lower

9. To ensure that other effects that we report below produce a Bayesian ratio of 3.0 or higher in favor of the alternative hypotheses, we also used JASP with default priors to run a Bayesian repeated measures ANOVA with task procedure (simple *vs.* complex span task) and category-set complexity (chunkable *vs.* non-chunkable) as repeated factors. We obtained a Bayes factor of $1.21e^{79}$ in favor of the alternative model (a model including a main effect of task procedure, a main effect of category-set complexity, and an interaction between the two), against the null model. Also, the Bayes factor for this alternative model against a model without the interaction component was $1.22e^{6}$. Other simpler models (one model including only an effect of task procedure, another model including only an effect of category-set complexity, or else another model including both effects without interaction) were all much better than the null model, but there was more evidence for the full model. Analysis based on model averaging (combining models that included a particular effect, either task procedure, category-set complexity, or their interaction) showed that changes from prior to posterior inclusion odds were all decisive (Bayes factor $> 10e^{2}$).

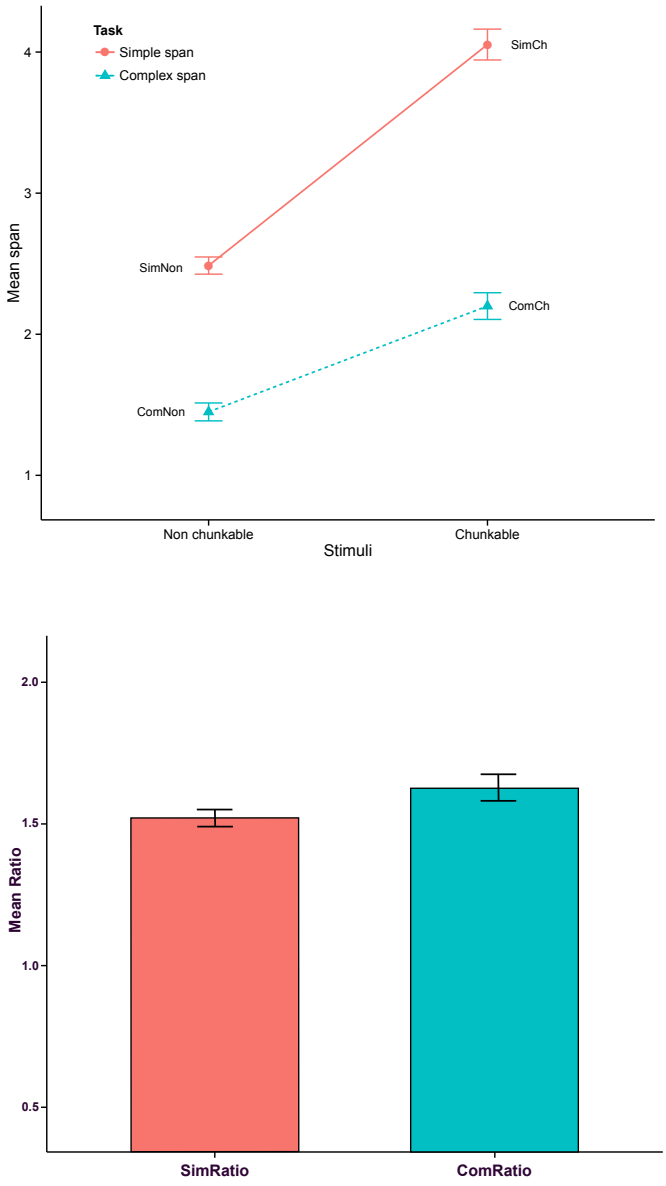


Figure 8 – Mean span for Experiment 1.

than what our hypothesis predicted, that is, lower than the lower bound of the 4 ± 1 capacity limit. For this experiment, we chose the simplest design, i.e., we ran only the *SimNon* condition of the preceding experiment.

Method

Participants

Sixty-five students enrolled at the University of Franche-Comté or the University Nice Sophia Antipolis, $M = 21.4$ years ($sd = 3.7$), volunteered to participate in this experiment.

Procedure

The procedure was similar to that of Experiment 1, except than only the *SimNon* condition was implemented. The only other difference was that the task was speeded up by using only two trials per stimulus length.

Scoring

A score of .5 was given for each perfectly correct serial report. For instance, a participant who correctly recalled two lists of one object, two lists of two objects, and only one list of three objects was assigned a span of 2.5.

Results

Figure 9 (left) shows the histogram of the span. A mean span of 2.46 ($sd = .63$) was found across participants, a value below the usual 4 ± 1 capacity limit. Table 3 shows that 2.46 was

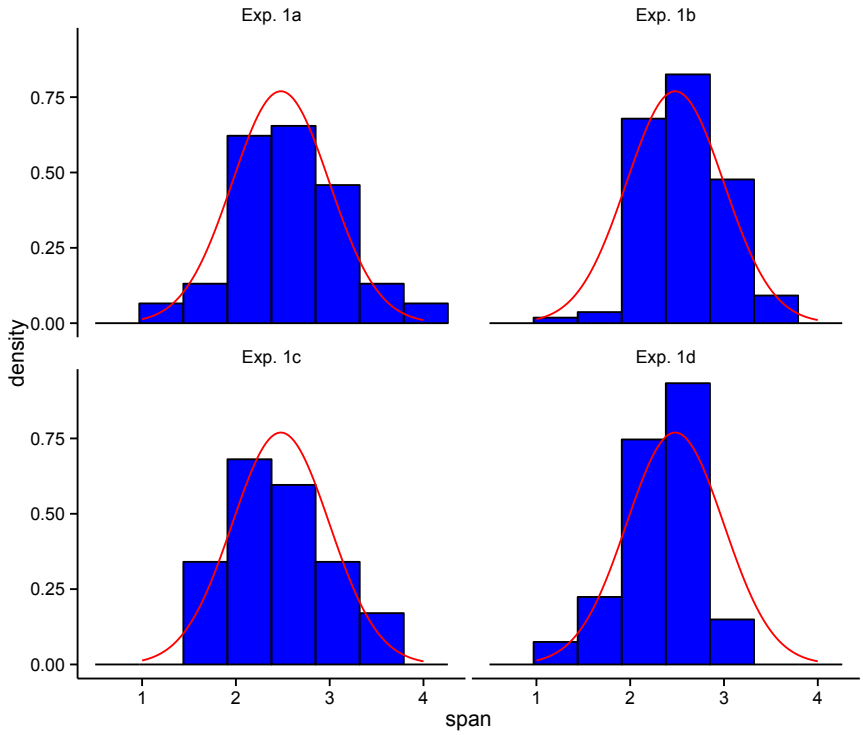


Figure 9 – Histograms of the the mean spans observed in Experiment 2A (top left), and for the second phase of Experiment 2B (top right), for the second phase of Experiment 2C (bottom left), and in Experiment 2D (bottom right). *Note.* The spans were computed using the all-or-nothing-unit-scoring method (i.e., one point was cumulated for each perfectly recalled sequence, and the total was divided by the number of repetition per list length.

Table 3 – *Mean Span for nonchunkable material and simple span task procedure, measured in Experiment 1, 2A, 2B, 2C, and 2D for adult participants, compared to an average span of 3 using one-sample t-tests. For Exp. 2A and 2D, the average span of the unique condition is reported. For Exp. 2B and 2C, the highest span measured across all conditions is reported to run a conservative test.*

Task	M (se)	Compared to an average span of 3
1	2.49 (.06)	$t(93) = -8.4, p < .001$
2A	2.46 (.08)	$t(64) = -6.9, p < .001$
2B	2.64 (.06)	$t(56) = -6, p < .001$
2C	2.4 (.13)	$t(24) = -4.7, p < .001$
2D	2.33 (.06)	$t(56) = -11.8, p < .001$

significantly below the lower limit of the 4 ± 1 capacity generally observed in complex span tasks.

Experiment 2B

The low capacity found in Experiment 2A can be explained in two ways: (1) the possibility of grouping any of the consecutive objects was severely limited by our manipulation, as expected, or (2) the experiment was too short (participants had only two trials per list length), so there may not have been enough time for the participants to become familiar with the task. Experiment 2B was run to test if a longer experiment could help participants become more familiar with our material by using, for instance, a greater number of trials. We conducted a second experiment, based on the same design as in Experiment 2A, except that we increased the duration of the experiment by manipulating three factors. (1) We increased the number of trials per length. (2) The task was administered twice in a row (to give a second chance to the participants to perform better). (3) We compared two groups by varying the stimulus display time (1 second *vs.* 2 second). The two-second condition was thought to allow for better encoding of the stimuli.

Method

Participants

One hundred and sixteen students enrolled at the Université Nice Sophia Antipolis, $M = 21$ years ($sd = 3.6$), volunteered to participate in this experiment.

Stimuli

The stimuli and their random combinations were similar to those used in Experiment 2A (see Figure 7).

Procedure

The procedure was like the one used in Experiment 2A, except that three variables were added to the design to offer the participants more opportunity to achieve a higher span. First, the participants were administered four trials per length instead of two. The experiment automatically stopped after four errors within a given list length. Second, memory performance was studied across two different stimulus-onset asynchronies (SOA; one second for 58 participants and two seconds for 57 participants). Third, the task was run twice in a row (the two sessions were labelled first phase and second phase).

Scoring

A score of .25 value was given for each perfectly correct serial report. For instance, a participant who correctly recalled four lists of one object, four lists of two objects, and only one list of three objects was assigned a span of 2.25.

Results

We performed a 2 (first phase *vs.* second phase) × 2 (1-sec *vs.* 2-sec) ANOVA using the mean span for each participant as the dependent variable. A significantly greater span was obtained in the second phase of the experiment ($M = 2.58, sd = .44$) than in the first phase ($M = 2.31, sd = .56$), $F(1, 114) = 29.96, p < .001, \eta_p^2 = .21$. Also, a significantly greater span was obtained using a two-second SOA ($M = 2.54, sd = .51$) than a one-second SOA ($M = 2.35, sd = .47$), $F(1, 114) = 5.6, p = .02, \eta_p^2 = .05$; there was no interaction with the phase factor.

During the first phase, analysis of the simple effects showed that the 2-sec SOA led to a higher span ($2.44, sd = .56$) than in the 1-sec SOA ($2.19, sd = .55$) ($t = -2.46, df = 113, p = .015$). However. SOA had no effect during the second phase of the experiment, where the spans were 2.52 (.42) and 2.64 (.45), respectively. The highest value obtained across the four experimental conditions, that is, 2.64 on the second phase for the 2-sec SOA (reported in Figure 9), which was the most favorable condition, was significantly below (Table 3) the lower limit of the 4 ± 1 capacity generally observed in complex span tasks.

Experiment 2C

This experiment tested whether the visual complexity of the Boolean stimuli or the relational complexity between the Boolean stimuli that caused the difficulty for the participants to recall more than three items as observed in the previous experiments. We chose to explore wether adults would encode the boolean stimuli differently than children. One reason is that the verbalization of nonverbal stimuli can impair recall (a process sometimes called verbal overshadowing; Melcher & Schooler, 2004). Given the hypothesis that adults can use a more complex reper-

toire of verbal strategies than children, we posit than one possibility for the difficulty of the participants in the first two experiments was the use of a verbal strategy that was detrimental to recalling the stimuli. Effectively, a verbal strategy can reveal inefficient in re-encoding the sequences of stimuli because the heterogeneity of the category sets makes the features uncorrelated, and thus numerous to encode.

Method

Participants

A total of 99 participants were split into four Age-groups; 7 year-olds ($M = 6.8, sd = .37, n = 25$), 9 year-olds ($M = 8.8, sd = .31, n = 22$), 11 year-olds ($M = 10.9, sd = .34, n = 27$), from a public school, and a group of young adults enrolled at the university Nice Sophia-Antipolis ($M = 22, sd = 1.76, n = 25$). The children were from middle-class families. All the children participated voluntarily, and their parents signed an informed consent form.

Procedure

The four Age-groups of Participants were tested on two visual span tasks. The order of administration of the tasks was the same for each participant, and was as follows: a simple span task based on the stimuli used by (Alvarez & Cavanagh, 2004); Immediate recall of a short story and a visual span task similar to Exp. 2A and B. The whole experiment lasted approximately 20 min.

Visual span task based on Alvarez & Cavanagh, 2004: One advantage in using the six categories of objects shown in Fig.10 is that Alvarez and Cavanagh (2004) have shown that capacity limit varies substantially across the six stimulus classes in the order shown in Fig.10, with a smaller

capacity for the shaded cubes and a greater capacity for colors. One difficulty for adapting this material to a simple span task is that each of the categories only contains six different items. Therefore, for each category, we used only one trial per list length up to $L = 6$. The stimuli used for $L = 1$ (for instance, a), $L = 2$ (for instance, bc), and $L = 3$ (for instance, def), were repeated for the last three trials: $L = 4$ (for instance, $bacd$), $L = 5$ (for instance, $ebdfa$), and $L = 6$ (for instance, $fabfed$). The lists were built for each participants in a pseudo random order to avoid the repetition of pairs across lengths (e.g.,we would avoid lists such as bd, abd, $fabd$, etc.) and the order of the categories were randomly ordered between participants. The participants were required to reconstruct the original stimulus order by clicking on the response screen of the L randomly placed objects.

Visual span task based on Boolean stimuli: The participants were then administered the simple span task using the Boolean stimuli of Exp. 2A and Exp. 2B. After a first phase, which followed the exact same procedure as phase 1 of Exp. 2B (display time: 1 sec.), participants had to complete a second phase, for which they were given instructions on how to memorize the stimuli. Half of them were asked to verbalize the task while the other half was invited to retain the stimuli visually (the children were instructed to take a mental photography of the objects).

Scoring

For the first task (using the stimuli of Alvarez & Cavanagh, 2004), 1 point was cumulated for each perfectly correct serial report and for the second task (using the Boolean stimuli), a .25 value was cumulated for each perfectly correct serial report. The spans were computed using the all-or-nothing-unit-scoring method. For instance, a participant who correctly recalled in the first task four lists of one object, three lists of two objects, and one list of three objects was granted a span of $(4 + 3 + 1)/4 = 2$.

Results

Effects of Age-group and Stimulus class

An ANOVA was performed, including the Stimulus class (alphabet, colors, objects, kanjis, cubes, polygons, boolean) as repeated factor and Age-group (7, 9, 11, and Adults), and using the mean span as the dependent variable. For the task using the boolean stimuli, we only took into account the first phase, because no instruction (verbalize or visualize) was given to the participants, as it was also the case in the tasks using the other stimulus classes. This analysis showed a main effect of Stimulus class on the span, $F(6, 588) = 96.77, p < .01, \eta_p^2 = .5$. For the post-hoc analyses, we chose to run a single pairwise comparison of the mean span between the boolean stimuli ($M = 1.83, se = .06$) and the polygons ($M = 2.68, se = .12$) because the polygons led to the lowest span among stimulus classes (cubes: $M = 2.78, se = .12$; kanji: $M = 3.22, se = .11$; objects: $M = 3.88, se = .12$; colors: $M = 4, se = .12$; alphabet: $M = 4.49, se = .13$). This comparison showed a significantly lower span ($t(98) = 7.39, p < .01$) for the boolean stimuli than for the polygons. Aside the expected effect of age on the span ($F(3, 95) = 62.46, p =< 01, \eta_p^2 = .66$), we also found an interaction between the two factors ($F(18, 570) = 1.86, p = .017, \eta_p^2 = .05$; this interaction is shown in Fig. 11), which globally represents greater differences between stimulus classes as age increases.

To follow the analyses conducted in the two previous experiments, we then restricted the analysis to the adults' span for the boolean stimuli, and again, the capacity was found significantly inferior to the reference span of 3 ($t(56) = -11.8, p < .001$, see Table 3).

Table 4 – *Mean Spans (and std errors) measured in Exp. 2C by Age-group by phase and by instruction type (verbalize vs visualize), for the boolean stimuli.*

Phase	Condition	Age-group			
		7 (n = 25)	9 (n = 22)	11 (n = 27)	22 (n=25)
First	none	1.28 (.1)	1.68 (.09)	2.01 (.12)	2.31 (.09)
Second	Verbalize	1.1 (.14)	1.72 (.11)	1.84 (.14)	2.31 (.17)
Second	Visualize	1.33 (.14)	1.79 (.14)	1.71 (.13)	2.5 (.2)

Effects of instructions on the span of Boolean stimuli

We performed an ANOVA with Age-groups (7, 9, 11, and Adults) and Instruction (verbalize vs. visualize) as between-subject factors and using the mean span as the dependent variable. This Analysis showed a significant effect of Age ($F(3, 98) = 21.05, p < .001, \eta_p^2 = .92$), no effect of the instructions ($F(1, 98) = .69, p = .41, \eta_p^2 = .01$), and a significant interaction ($F(7, 98) = 9.2, p < .001, \eta_p^2 = .42$). The interaction is due to the slight variable effect of the instructions among age groups, that is shown in Table 4. However, when we conducted a series of t-tests to compare the mean spans between the instructions Verbalize and Visualize for each age group (7, 9, 11, and Adults), none was statistically significant ($t(23) = -1.12, p = .27$; $t(20) = -.38, p = .71$; $t(25) = .66, p = .52$, $t(22) = -.74, p = .47$, respectively). A last analysis of variance showed a main effect of the repeated factor Phase (first vs. second) on the span ($F(1, 95) = 0.32$), an effect of age ($F(3, 95) = 26.2, p < .001, \eta_p^2 = .45$), and no significant interaction ($F(3, 95) = 2.02$). The slight positive effect of the second phase ($M = 1.82, se = .05$) in comparison to the first phase ($M = 1.79, se = .05$) could be due to either a gain of practice or the fact of being instructed by one of the two strategies.

Discussion

The aim of this study was to examine (1) developmental effects in memorizing sequences of boolean stimuli chosen among heterogeneous category (in comparison to other stimulus classes)

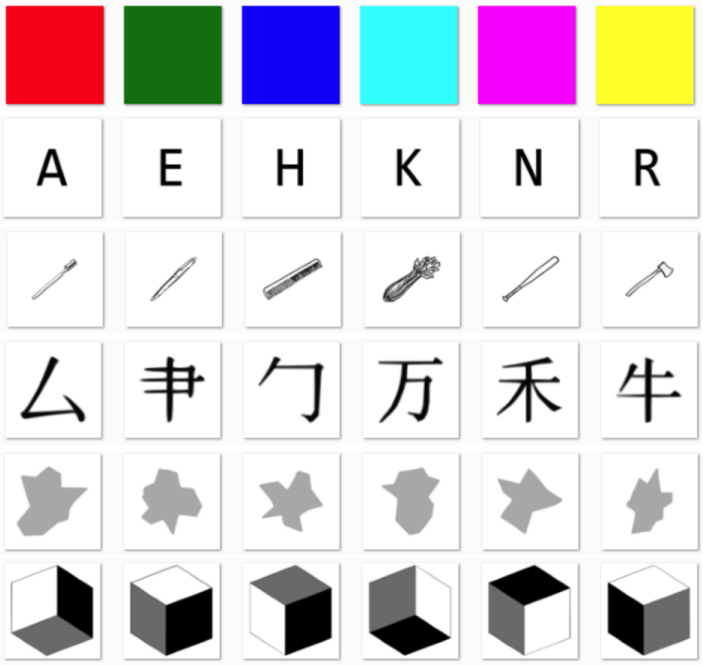


Figure 10 – Stimuli used by Alvarez and Cavanagh (2004), and in Exp. 2C. From top to bottom: colors, alphabet, objects, kanjis, polygons, cubes.

and (2) detrimental effects of verbalization on recalling the Boolean stimuli and (3) an interaction between these two effects. Our results show that the instruction given to the participants to verbalize the stimuli had no global effect on recall performance. Also, the slight interaction found between the verbalize vs visualize factor and Age seems not due to a systematic trend such as an increasing detrimental effect of verbalization with Age, but rather to a series of irregular variations of the verbalize vs visualize factor across group ages. The only solid interaction we found was between the stimulus class factor and Age, with greater differences between stimulus classes as Age increases. Overall, and more importantly, our boolean stimuli still appear to be the most difficult to memorize in comparison to the other classes of visual stimuli that were studied by Alvarez and Cavanagh (2004), and this difficulty cannot be easily modified by a verbal or a visual strategy.

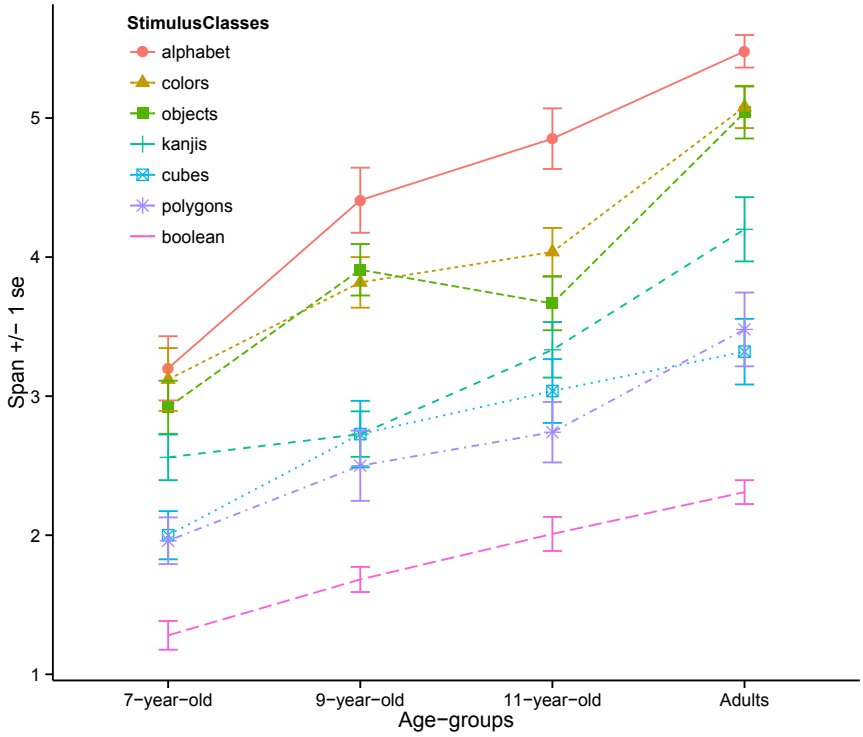


Figure 11 – Mean spans for Experiment 2C. Note: the results plotted in the boolean condition represent the span measured during the first phase of the task, which involved no specific instruction.

Experiment 2D

The aim of this experiment was to adapt Exp. 2B to a change detection paradigm. The reason is that brief presentation of all items at once can reduce verbal strategies that we still suspect to account for the difficulties to encode our boolean stimuli. The change detection paradigm is also known to reduce rehearsal and chunking mechanisms. We thus chose to display the objects all at once during 250ms instead of presenting them serially like in the preceding experiments.

Method

Participants

Fifty seven students enrolled at the University of Franche-Comté, $M = 19.7$ years old ($sd = 1.2$), volunteered to participate in this experiment.

Procedure

As in Exp. 2B, the number of items per list progressively increased (from one to eight objects) until supraspan conditions were reached. All the objects of a list were displayed at once in random positions, during 250ms. There were four different pseudorandom lists of items per list length and the task automatically stopped after four consecutive errors per list length, to follow the stop criterion adapted in Exp. 2B and 2C.

Scoring

Scoring was similar to Exp. 2B and 2C, but we also computed a different span (hereafter: k) using the hits and false alarms (FA) to have a more refined indicator of the span. For each list

length, we computed the span using $k = n\ (hits - FA)$, using n = number of displayed objects. We then computed the max(k) among list lengths for each participant.

Results and Discussion

In this Experiment, the participants' mean span was 2.33 on average ($se = .06$) (see Table 3 and bottom right of Figure 9). Again, this capacity was significantly inferior to 3 (Table 3). Using the second scoring method, we found $k = 3.25$ ($se = .09$), which was found significantly superior to 3 ($t(56) = 2.8, p < .01$), but significantly inferior to 4 ($t(56) = -8.6, p < .01$).

General Discussion

The initial aim of the current study was to show that performance in a simple span task can be equal to that of a complex span task (i.e., around 4 ± 1) as long as no regularity is present in the stimulus lists. It is rarely the case that no information can be grouped together when one uses alphanumerical stimulus lists (Jones & Macken, 2015) or visuo-spatial tasks such as the Corsi (Gmeindl, Walsh, & Courtney, 2011 ; Haladjian & Mathy, 2015), the most common type of material for simple span tasks. Therefore, we used nonverbal objects for which we could facilitate or impede grouping by varying the amount of regularity or compressibility in the sequence.

Firstly and surprisingly, we found no fundamental difference between the simple and complex span tasks regarding the ability to recode information. We showed that a chunking process can operate in complex span tasks in a manner comparable to simple span task inasmuch as we found a ratio of about 1.5 in the compressible:incompressible spans in both simple and complex span procedures.

This could mean that concurrent tasks simply prevent participants from refreshing information, and that chunking and related compression processes can occur even in complex span tasks and therefore may be automatic. We therefore suggest that encoding to-be-remembered items in a complex span task still includes a compression process. Although the present study was conducted primarily to design a new simple span task, further experiments focusing on the complex span-task procedure could test whether more severely limited timing (such as reading numbers aloud at a rate so fast that there is no time in between) could prevent data compression. Experiment 1 thus disconfirmed that complex spans cannot be subject to a compressibility effect. Compression seems more ubiquitous than predicted, so compressibility is not the reason for the difference between simple and complex span tasks. Overall, based on the ratio of the number of encoded objects in the two conditions, we found that it is similarity between the two types of spans that prevails. This conclusion is consistent with previous findings using a correlational approach that distinction between simple and complex span tasks can be questioned. For instance, Unsworth and Engle (2007b) published one comprehensive review acknowledging the findings of Colom, Rebollo, et al. (2006) and Colom, Shih, Flores-Mendoza, and Quiroga (2006), that simple and complex span tasks may rely on the same basic processing resources and cannot be properly distinguished (see also Martínez et al., 2011).

Secondly, using incompressible sequences designed to hinder rapid re-encoding of information, we showed that the average span was less than three objects using a simple span task. This surprisingly low value is below the reported 4 ± 1 capacity of working memory most often found in the literature (Cowan, 2001). However, our result matches the mean of 2.38 found for the condition with the highest cognitive load condition in a complex span task reported by Vergauwe, Barrouillet, and Camos (2010). We contend that our new simple span task can offer a purer measure of storage capacity than any of the earlier simple span tasks that have used

verbal or visuo-spatial material.

Although it may seem troubling to obtain such an unexpectedly low span, such a concern would be based on the assumption that the limit should apply to a certain fixed number of objects, with each object including all of its features bound together (Luck & Vogel, 1997). However, there is more recent evidence that features must be held in mind separately (Cowan, Blume, & Saults, 2013 ; Hardman & Cowan, 2015 ; Oberauer & Eichenberger, 2013 ; B. Wang, Cao, Theeuwes, Olivers, & Wang, 2016). Our results show that capacity is still related to a fixed number of objects in the 4 ± 1 range, but not necessarily with all of their features intact. It might be possible that participants retained about four objects, but some objects might not be perfectly encoded. Future tests using single-item probes in the test display should help address this question. Because our scoring system was based on the all-or-nothing procedure (i.e., a single unencoded feature would cause failure), it is not that surprising to find that a smaller capacity limit was found using k-object lists that included $3k$ features. However, it is still notable that the number of relatively common objects (i.e., our objects were not complex polygons or kanjis, as in Alvarez and Cavanagh (2004), for example), that could barely be held in memory without any kind of disruption or limitation of attention was only two.

Quatrième partie

Mesure de la capacité de la mémoire de travail par une tâche d'empan de chunks et relation avec l'intelligence

Compression in Working Memory and its Relationship to Fluid Intelligence

Mustapha Chekaf

Université de Franche-Comté

Nicolas Gauvrit

CHArt Lab (PARIS-reasoning)

Alessandro Guida

Université Rennes II

Fabien Mathy [10]

Université Nice Sophia Antipolis

10. Author's note : Fabien Mathy, Département de Psychologie, Université Nice Sophia Antipolis, Laboratoire BCL : Bases, Corpus, langage - UMR 7320, Campus SJA3, 24 avenue des diables bleus, 06357 Nice CEDEX 4. Email : fabien.mathy@unice.fr. This research was supported in part by a grant from the Région de Franche-Comté AAP2013 awarded to Fabien Mathy and Mustapha Chekaf. We are also grateful Yvonnick Noël and to the Attention & Working Memory Lab at Georgia Tech for helpful discussion, and particularly to Tyler Harrisson who suggested two of the structural equation models. Authors' contribution : FM initiated the study and formulated the hypotheses ; MC and FM conceived and designed the experiments ; MC run the experiments ; MC, NG, and FM analyzed the data ; NG computed the algorithmic complexity ; MC, NG, AG, and FM wrote the paper.

Abstract:

Working memory has been shown to be strongly related to fluid intelligence, but our goal is to go further to shed light on the process of information compression in working memory as a determining factor of fluid intelligence. Our main hypothesis was that compression in working memory is an excellent indicator for studying the relationship between working-memory capacity and fluid intelligence, because both depend on the optimization of storage capacity. Compressibility of memoranda was estimated using an algorithmic complexity metric. The results showed that compressibility can be used to predict working-memory performance, and that fluid intelligence is well predicted by the ability to compress information. We conclude that the ability to compress information in working memory is the reason why both manipulation and retention of information are linked to intelligence. This result offers a new conception of intelligence based on the idea that compression and intelligence are equivalent problems.

Compression in Working Memory and its Relationship to Fluid Intelligence

Although there is no doubt that working memory (WM) is closely related to fluid intelligence (Kane, Hambrick, & Conway, 2005), there is no clear-cut answer as to why manipulation and retention of information are responsible for variations in people's fluid intelligence. The main account in describing this relationship that has been given so far is based on memory capacity. For instance, the role of capacity becomes increasingly greater with the more difficult items of Raven's Progressive Matrices (Little, Lewandowsky, & Craig, 2014). Although WM is known to play an indisputable role in intelligent behaviors, our goal is to go further in order to shed light on the process of information compression in WM as a determining factor of intelligence. Compression of information, that is, the capacity to recode information into a more compact representation may be a key mechanism in accounting for the dual impact of the manipulation and retention of information on intelligence, particularly when people are dealing with new problems in the typical tests used to measure intelligence (such as Raven's Matrices).

One observation is that WM compression processes show up in memory-span variance (Chekaf et al., 2016). As modeled by Mathy and Feldman, the 7 ± 2 estimation first found by Miller (1956) may correspond to an overestimation due to compression [11] of the true capacity of immediate memory (which is now most often estimated as being about 4 ± 1; e.g., Broadbent, 1975 ; Cowan, 2001 ; Luck & Vogel, 1997). The possibility that the span ranges between 4 and 7 is made crucial by what we call the short-term memory/WM paradox. Conceptually, short-term memory (STM) is thought to be primarily involved in storage, but since storage is a relatively simple task, items can be manipulated and potentially compressed. This may explain why the span can go as high as seven items on the digit-span task. Conversely, while WM is primarily involved in both the manipulation and storage of information, the capacity

11. Another account in terms of the participant's linguistic experience was given by Jones and Macken (2015).

to manipulate information is generally overlooked and the focus is mainly on storage (for an exception, see Unsworth et al., 2009). The present paper proposes and develops the idea that the concept of information compression can help eliminate the opposition between STM and WM and may be useful in accounting for intelligence in a more comprehensive way.

Compression can potentially account for intelligence (for similar ideas in artificial intelligence, see Baum, 2004; or Hutter, 2005) because many complex mental activities still fit well into rather low WM storage capacity. Our goal here is to show that the ability to compress information (see Brady et al., 2009) is a good predictor of WM performance and is thereby a good candidate for predicting intelligence: more compression simply means more available resources for processing other information, which can potentially have an impact on reasoning.

To study the ability to compress information, we developed a task based on the SIMON® framework, a classic memory game from the 80s that can be played on a device with four colored buttons (red, green, yellow, blue). The game consists of reproducing a sequence of colors by pressing the corresponding buttons. The SIMON span task was developed on the reasoning that sequences of colors contain regularities that can be quantified to approximate a compression process. To estimate the compressibility our participants could be offered within thousands of different color sequences in the SIMON task, we devised a compressibility metric that provides an estimate of every possible rapid grouping process that can take place in WM. Experiment 1 was very liberal in terms of randomly choosing the sequences of colors, which lead us to develop a specific estimation of memory capacity, whereas Experiment 2 used similar sequences across participants that allowed us to use a more standard scoring method for computing a memory span. Although there are already reports studying the relationship between the components of the WM system and intelligence at the latent variable level (see Chuderski, Taraday, Nęcka, & Smoleń, 2012 ; Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004 ; Krumm et

al., 2009 ; Martínez et al., 2011 for examples), the present study combines experimental and correlational approaches because of the exploratory nature of the new task. The results showed that this metric can indeed be used to predict WM performance, and that intelligence is well predicted by the ability to compress information. We concluded that the ability to compress information in WM is the reason why both the manipulation and the retention of information are linked to intelligence.

Experiment 1

Experiment 1 develops a chunking memory span task based on an algorithmic complexity metric for short strings and study how it relates to intelligence. This metric enabled us to estimate a capacity in STM, in which storage and processing are combined, by measuring the most complex sequence that could be remembered by each participant. The sequences were drawn at random in order to study complexity effects independently from participant's individual performance, but with the main goal of having a maximum number of different sequences to fit a global performance function. We predicted that performance would decrease with complexity because complexity hinders the capacity to compress information. It was also predicted that performance on our task would better correlate with other span tasks that solicit a combination of storage and processing, but particularly when processing is dedicated to the stored items such as a memory updating task (i.e., this is not the case in complex span tasks).

One interesting battery developed by Lewandowsky, Oberauer, Yang, and Ecker (2010), the *Working Memory Capacity Battery (WMCB)* allowed us to test which of the *storage* − *processing* combinations best predicts IQ, and whether proximity with our task could be predicted by the *storage* − *processing* combination specific to each task. For the two complex span tasks of the WMCB, processing is directed away from the stored items. For the spatial

span task of the WMCB, the processing component is left uncontrolled and in this case, the processing component is not fully involved but only moderately dedicated to the stored items. The memory updating task of the WMCB can however be regarded as a difficult task in which processing is fully dedicated to the manipulation of the stored items. Again, the processing component is not fully involved in this task because the number of items updated is linearly dependent on the number of stored items, meaning that if the participant's storage capacity is low, processing of the items is also directly limited. Our hypothesis was that our task draws on a clearer *storage* × *processing* combination in which the processing component can be used to optimize storage capacity, and in this particular case, the number of items that are processed sometimes exceeds the number of groups that are stored. In this latter case, *storage* capability (in terms of the number of chunks that can be stored) can be limited, but the number of items that can be packed into the chunks is less constrained. Accordingly, our task is expected to better correlate with the simple span task and the memory updating task than with the complex span tasks.

Method

Our goal was to devise a new span task enabling us to measure the human ability to compress information in WM. Concurrent measures of WM were administered for comparison with the new span task and its predictions of fluid intelligence.

Participants

One hundred and eighty-three students enrolled at the University of Franche-Comté in Besançon, France (M_{age} = 21; SD = 2.8) volunteered to participate in the experiment and received course credit in exchange for their participation.

Procedure

Depending on their availability and their need for course credit, the volunteers took one, two, or three tests: our adaptation of the game SIMON®, (hereafter called the SIMON span task), the Working Memory Capacity Battery (WMCB), and Raven's Advanced Progressive Matrices (APM) (here called the SIMON span task), the Working Memory Capacity Battery (WMCB) and the Raven's Advanced Progressive Matrices (APM) (Raven, 1962). All volunteers were administered the SIMON span task; some also took the WMCB ($N = 27$), Raven's matrices ($N = 26$), or both [12] ($N = 85$). In all cases, tests were administered in the following order: (1) SIMON span task, (2) Raven, (3) WMCB.

SIMON span task

Each trial began with a fixation cross in the center of the screen (1000 ms). The to-be-memorized sequence consisting of a series of colored squares appearing one after the other was then displayed (see Figure 12). Next, in the recall phase, four colored buttons were displayed and participants could click on them to recall the whole sequence they had memorized and then validate their response. After each recall phase, feedback ('perfect' or 'not exactly') was displayed according to the accuracy of the response.

Participants were administered either a spatial version or a nonspatial version of the task. In the spatial version (N = 106), the colored squares were displayed briefly lit up one at a time, in different locations on the screen, to show the to-be-remembered sequence, as in the original game. To discourage spatial encoding, the colors were randomly assigned to the locations for

12. Although there are new guidelines for sample size requirements (Wolf, Harrington, Clark, & Miller, 2013), we simply followed the rule-of-thumb of a minimum sample size of 100 participants (those administered with the Raven's matrices) for running structural equation models including six observed variables.

each new trial. In the nonspatial version (N = 77), the colors were displayed one after another in the center of the screen in order to avoid any visuo-spatial encoding strategy. To further discourage spatial strategies in both versions, the colors were randomly assigned to the buttons on the response screen for each trial, which resulted in the colors never being in the same locations across trials. The reason for avoiding spatial encoding in both versions was that our metric was developed for detecting regularities in one-dimensional sequences of symbols, not two-dimensional patterns. Given that our results indicated no significant difference between these two conditions, the two data sets were combined in the analyses.

Each session consisted of a single block of 50 sequences varying in length (from one to ten) and in the number of possible colors (from one to four). New sequences were generated for each participant, with random colors and orders, so as to avoid the presentation of the items in ascending length (two items, then three, then four, etc.). We chose to generate random sequences and to measure their complexity *a posteriori* (as described below). A total of 9150 sequences (183 subjects $\times$ 50 sequences) were presented (average length = 6.24), each session lasted 25 min on average.

Working Memory Capacity Battery (WMCB)

Lewandowsky et al. (2010) designed this battery for assessing WM capacity by means of four tasks: an updating task (memory updating, MU), two span tasks (operation and sentence span, OS and SS), and a spatial span task (spatial short-term memory, SSTM). It was developed using MATLAB (MathWorks Ltd.) and the Psychophysics Toolbox (Brainard, 1997 ; Pelli, 1997). On each trial of MU, our participants were required to store a series of numbers in memory (from 3 to 5) that were displayed in separate frames, one after another, for one second each. Next, only the empty frames remained on the screen, and this was followed by a sequence of several cues

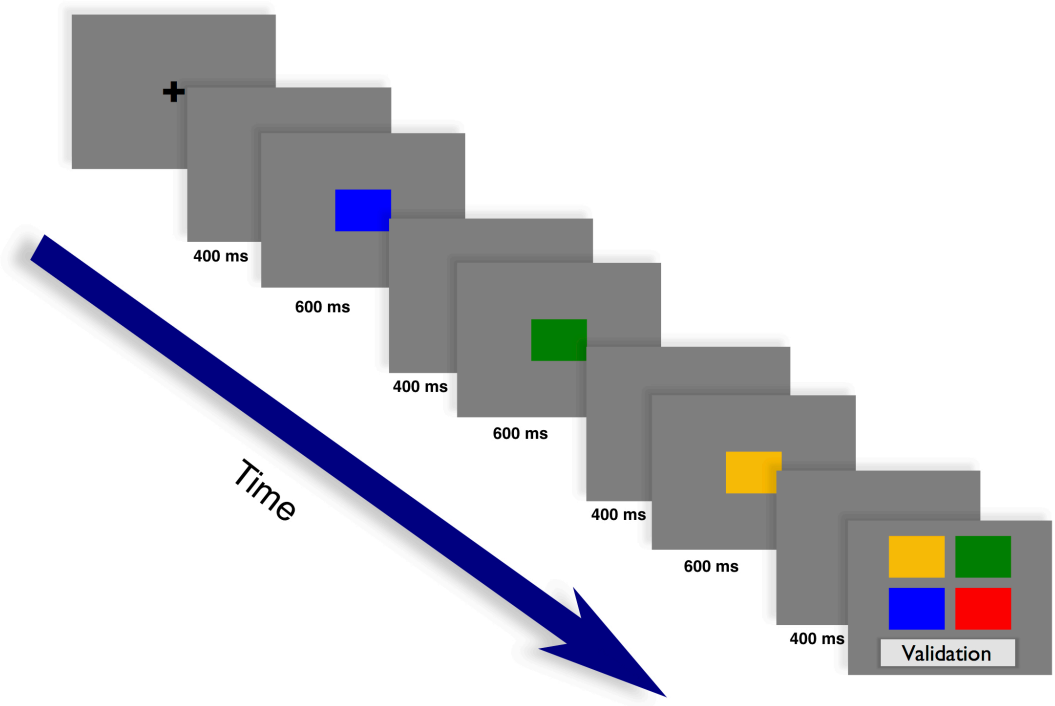


Figure 12 – Example of a sequence of three colors for the memory span task adapted from the SIMON® game.

corresponding to arithmetic operations such as "+3" or "−5" displayed in the respective frames one at a time in random frame. The participants had to apply these operations to update the memorized sequence of digits. In both OS and SS, a complex span-task paradigm was used in such a way that the participants saw an alternating sequence of to-be-remembered consonants and to-be-judged propositions. The judgments pertained to the correctness of the equations in the OS task or the meaningfulness of the sentences in the SS task. The participants were required to memorize the whole sequence of consonants for immediate serial recall. In SSTM, the participants were required to remember the location of dots in a 10×10 grid. The dots were presented one by one in random cells. Once the sequence was completed, the participants were cued to reproduce the pattern of dots using a mouse. They were told that the exact position of the dots or the order was irrelevant; the important thing was to remember the pattern made by the spatial relations between the dots. The four tasks together took 45 minutes on average.

Raven's Advanced Progressive Matrices

After a practice session using the 12 matrices in Set 1, the participants were tested on the matrices in Set 2 (36 matrices in all), during a timed session averaging 40 minutes. They were told to select the correct missing cell of a matrix from a set of eight choices. Correct completion of a matrix was scored one point, so the range of possible raw scores was 0-36.

Results

Effects of sequence length and number of colors per sequence

The effects of sequence length and the number of color per sequence are indicative of memory capacity in terms of the number of items recalled, and give us an initial idea as to whether the repetition of colors generated interference during the recall process (which would predict lower performance) or permitted recoding of the sequences (which would predict better performance). First, we conducted a repeated-measures ANOVA with sequence length as a repeated measure, and the mean proportion of perfectly recalled sequences for each participant (i.e., proportion of trials in which all items in a sequence were correctly recalled [13]) as the dependant variable. Performance varied across lengths, $F(9, 177) = 234.2, p < .001, \eta_p^2 = .9$, decreasing regularly as a function of length. Figure 13 shows the performance curve based on the aggregated data, by participant and length, and also as a function of the number of different colors in the sequence. The three performance curves resemble an S-shaped functions, as in previous studies (e.g., Crannell & Parrish, 1957 ; Mathy & Varré, 2013).

Next, we focused on a subset of the data in order to study the interaction between length and number of colors. We selected trials for which sequence length exceeded three items and we conducted a 7 (4, 5, 6, 7, 8, 9, or 10 items) $\times$3 (2, 3, or 4 different colors within a sequence) repeated-measures ANOVA. There was still a significant main effect of the length factor (Ms = .91, .82, .64, .49, .32, .2 and .13, and SEs = .01, .01, .02, .02, .02, .01 and .01 for the seven lengths, respectively), $F(6, 1092) = 560.8, p < .01, \eta_p^2 = .75$. The main effect of the number of colors per sequence was also significant (Ms = .6, .47 and .42; SEs = .01, .01 and .01),

13. Because repetitions occur within sequences, the proportion of correctly recalled items was not computed, insofar as it unfortunately involves more complex scoring methods based on alignment algorithms (Mathy & Varré, 2013), not yet deemed to be standardized scoring procedures.

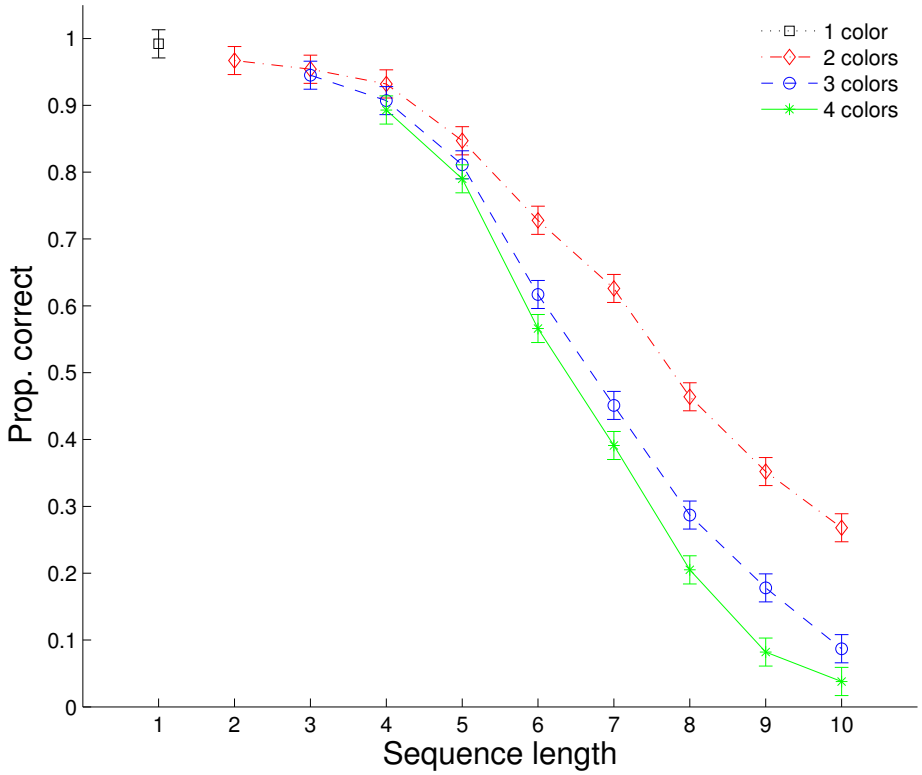


Figure 13 – Proportion of perfectly recalled color sequences, by sequence length and number of different colors within a sequence. Note: Error bars show ± one standard error.

$F(2, 364) = 132.4, p < .01, \eta_p^2 = .42$. Post-hoc analysis (with Bonferroni corrections for the pairwise comparisons) yielded a systematic, significant decrease between each length condition and between each number-of-colors condition. Finally, there was a significant interaction between length and number of colors, $F(12, 2184) = 5.8, p < .01, \eta_p^2 = .03$: the length effect increased with the number of colors. Although this is a coarse estimation of recoding, these results show that memorization was facilitated when sequences made of a fewer number of colors were used, and particularly when the sequences were longer.

Effect of compressibility

A long standing problem in this field is that short sequences of symbols (i.e., less than a hundred, as in the present study) are difficult to compress. The problem is that for most compression algorithms, rewriting *blue-blue-blue-red* to take into account the *blue-blue-blue* regularity can take more space than the original sequence. Thanks to recent breakthroughs, it is now possible to obtain a reliable estimation of the algorithmic complexity of short strings (Delahaye & Zenil, 2012 ; Soler-Toscano et al., 2013, 2014). The method has already been used in psychology, in domains other than working memory or intelligence (e.g., Gauvrit, Soler-Toscano, & Zenil, 2014 ; Kempe et al., 2015 ; Dieguez et al., 2015) and it has now been implemented as an R-package named ACSS (Algorithmic Complexity for Short Strings; Gauvrit et al., 2015). In this paper we hypothesize that this method is a simple approximation of intelligence.

To confirm the reliability of our complexity measure (which again, is inversely related to compressibility), we split each participant's sequences into two groups of equal complexity. This split-half method simulated a situation in which the participants were taking two equivalent tests. We obtained adequate evidence of reliability between the two groups of sequences ($r = .63, p < .001$); Spearman-Brown coefficient = .77).

Table 5 – *Stepwise forward selection of several models based on criterion BIC. At each stage, for a given model (where model ∼ v means that the base model is computed as a function of v to predict correct performance), the variables are listed in increasing BIC order after other variables were introduced. The final model includes complexity and length, but no interaction term. For instance, in the last model computed as a function of complexity and length, adding the interaction term increased the BIC, so the simplest model was chosen.*

Base model	Variable included	Deviance	BIC
Intercept only	Complexity	8093.7	8111.9
	Length	8462.3	8480.5
	None (intercept only)	12477.5	12486.6
∼Complexity	length	7805.2	7832.6
	None (complexity only)	8093.7	8111.9
∼Complexity & Length	None (complexity & length only)	7805.2	7832.6
	Interaction	7800.0	7836.5

A simple initial analysis targeting both the effect of the number of colors and the effect of complexity on accuracy (i.e., the sequence is perfectly recalled) showed that complexity ($\beta = -.62$) took precedence over the number of colors ($\beta = -.04$) in a multiple linear regression analysis when all 9150 trials were included. To further investigate the combined effects of complexity and list-length on recall in greater detail, we used a logistic regression approach. Selection of a stepwise forward model based on criterion BIC suggested dropping the interaction term (see Table 5). This model revealed a significant negative effect of complexity ($z(9147) = -23.84, p < .001$, standardized coefficient $= -5.70$; non standardized $= -.69$), as shown in Figure 14, and a significant positive effect of length ($z(9147) = 16.27, p < .001$, standardized coefficient $= 3.74$; non standardized $= 1.46$). Although length had a detrimental effect on recall, this effect was largely counteracted by the detrimental effect of complexity, meaning that long, simple strings were easier to recall than shorter, more complex ones. In other words, the complexity effect was stronger than the length effect (Table 5).

Figure 15 shows how performance decreased as a function of complexity in each length condition. The decreasing linear trend was significant for lengths 4, 6, 7, 8, 9, and 10 ($r = -.34, p < .001$;

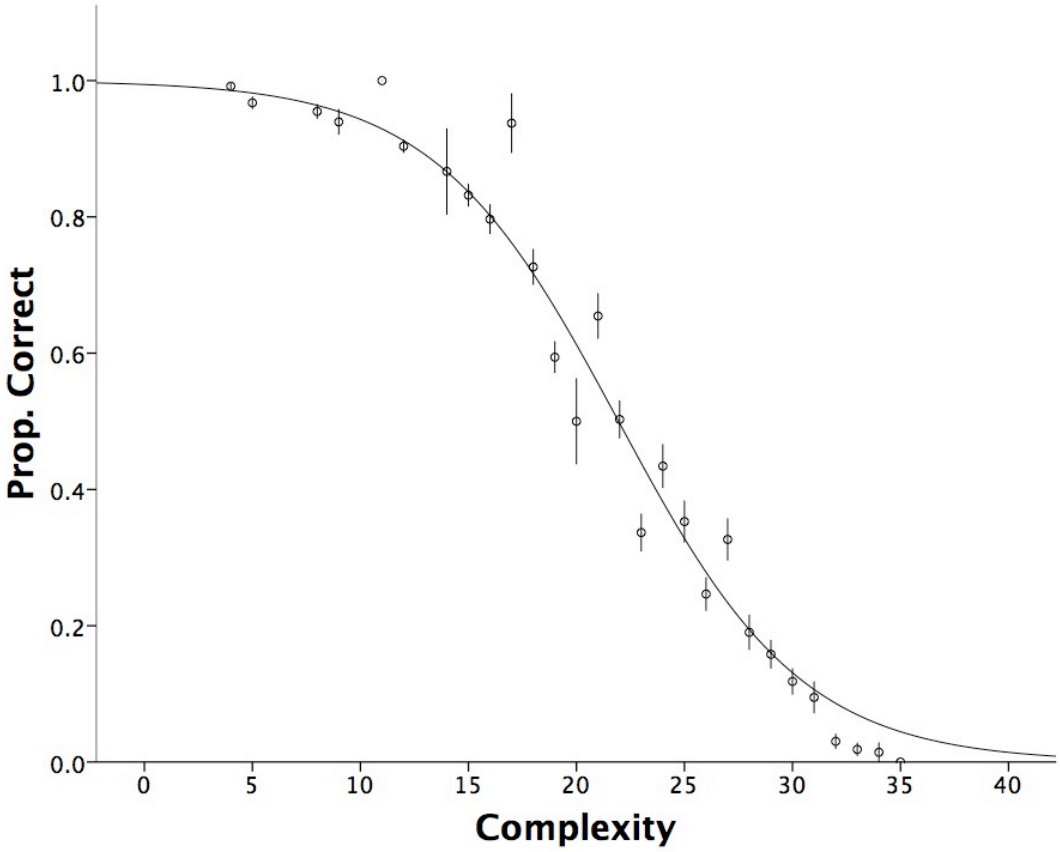


Figure 14 – Proportion of perfectly recalled sequences of colors as a function of complexity. Note: Error bars shows ± one standard error.

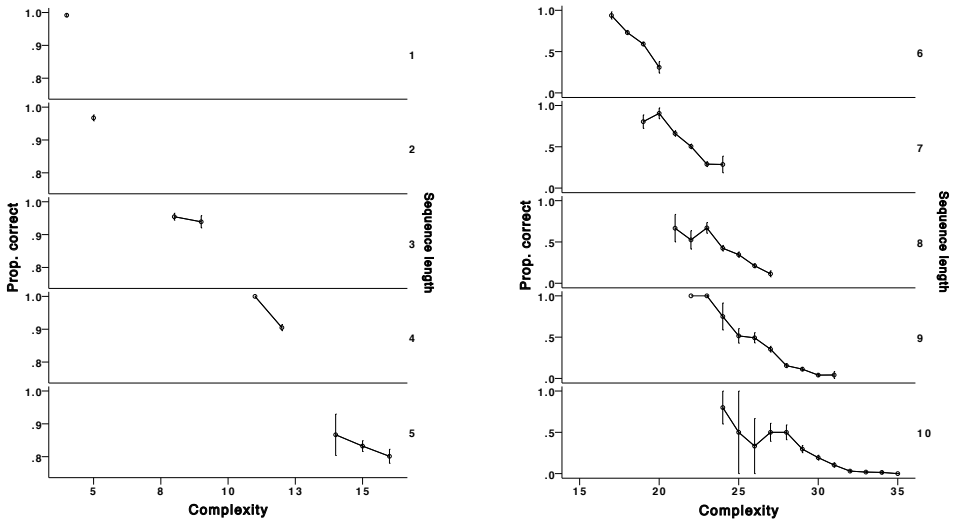


Figure 15 – Proportion of correct recall as a function of complexity, by sequence length. Error bars are ± one standard error.

$r = -.39, p < .001$; $r = -.41, p < .001$; $r = -.35, p < .001$; $r = -.45, p < .001$; and $r = -.43, p < .001$, respectively), which shows that the memory process could indeed be optimized for the less complex sequences.

One interesting result of our complexity measure was revealed in the comparison between complexity of the stimulus and response sequences. When the participants failed to recall the right sequence, they tended to produce a simpler string (in terms of algorithmic complexity). The mean complexity of the stimulus sequence was 25.3, but the mean complexity of the responses was only 23.3 across all trials ($t(9149) = 26.24, p < .001$, Cohen's paired $d = .27$).

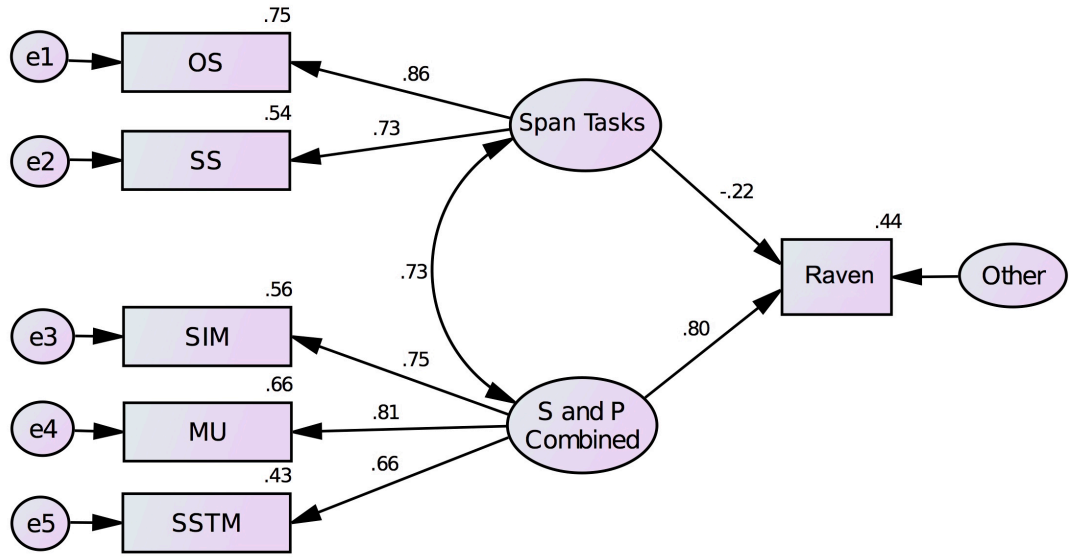
Following the regression on the data shown in Figure 15, we computed a logistic regression

for each subject to find the critical decrease in performance that occurred half-way down the logistic curve (i.e., the inflection point). The reason for this was that the participants were not administered the same sequences. This inflection point is thus based on complexity, and simply indicates that the participants failed on sequences more than 50% of the time when the complexity level was above the inflection point. To estimate the relationship between each participant's inflection point and his/her IQ, the data were input into confirmatory factor analysis (CFA) using IBM SPSS AMOS 21. A latent variable representing a construct in which *storage* and *processing* were separated during the task, and another latent variable representing a construct in which the two processes functioned together, were sufficient to describe performance. The fit of the model shown in Figure 16a was excellent ($\chi 2(7) = 2.85, p = .90$; comparative fit index CFI = 1.0; root mean squared of approximation RMSEA = 0.0; root-mean square residual RMR = .002; AIC and BIC criteria were both lower than in a saturated model with all variables correlated to each other, and than in an independence model with no variables correlated), with one caveat: the data failed to fit the recommended conditions for computing a CFA (see A. R. Conway et al., 2002 ; T. Wang, Ren, Li, & Schweizer, 2015) because one factor was defined by only two indicators, and also because the correlation between the two factors revealed collinearity. These results suggest that Raven's matrices are better predicted by the construct in which *storage* and *processing* are combined ($r = .64$, corresponding to 41% of the shared variance, instead of $r = .36$ when separated). The combined construct can be assessed in the present study by our SIMON span task, a memory updating task, and a simple span task. To make sure that the absence of SSTM would not weaken the predictions too much (since it is the only task that is mainly spatial), we tested a similar model that did not include the SSTM task (Figure 16b). In this case, we observed higher regression weights for Raven's matrices (.85 instead of .80, $-.26$ instead of $-.22$, and a percentage of shared variance with Raven's matrices of 43% instead of 41%; ($\chi 2(3) = 1.22, p = .75$). Figure 16c shows that when a factor already

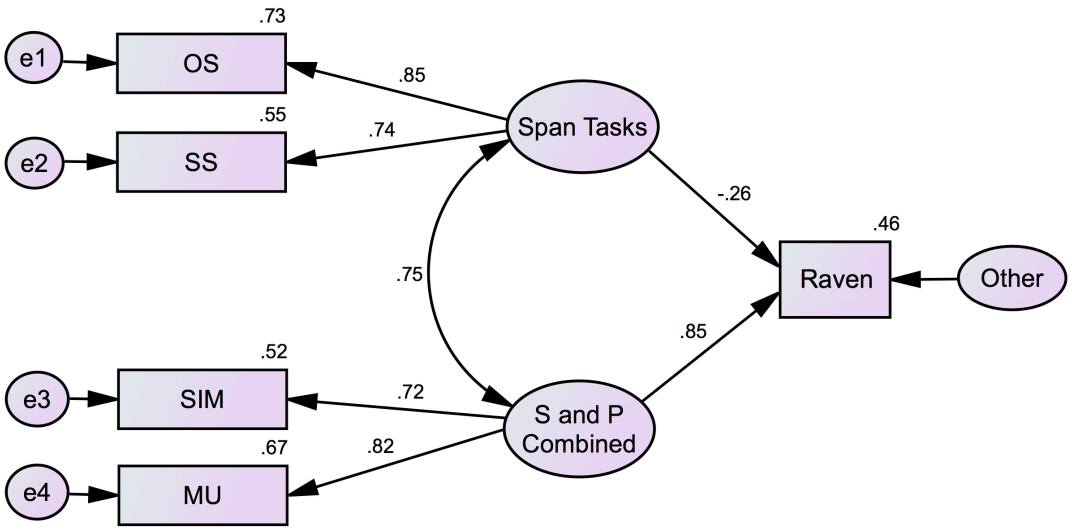
reflects all of the span tasks, a second factor reflecting only the tasks for which *storage* and *processing* are assumed to be combined still account independently for Raven performance. The substraction of the less restrictive model (Figure 16a) from the more restrictive model (Figure 16d) showed that forcing the two paths toward Raven's matrices to be equal significantly reduced the model's ability to fit the data ($\chi 2(9-7=2) = 54.95-2.85 = 52.1$), meaning that the .80 loading in Model A can be considered significantly greater than $-.22$.

Discussion

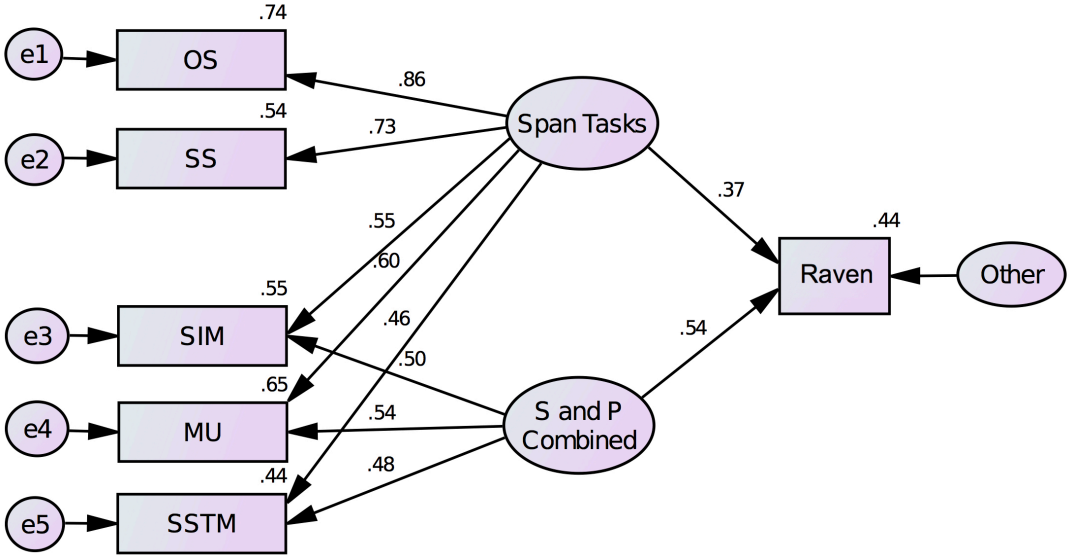
The objective of the first experiment was to investigate a span task integrating a chunking factor in order to manipulate the *storage* $\times$ *processing* interaction process. Our experimental setup was developed to allow the participants to chunk the to-be-remembered material in order to optimize storage. The chunking factor was estimated using a compressibility metric, to follow up on a previous conceptualization that a chunking process can be viewed as producing a maximally compressed representation (Mathy & Feldman, 2012). Using an algorithmic complexity metric for short strings, we acknowledge the difficulty in representing the chunking process precisely, sequence by sequence, and we preferred to adopt a more general approach stipulating that any regularity should be found and recoded by participants, particularly those with the greatest processing capacity. The hypothetical *storage* $\times$ *processing* construct in which processing is fully dedicated to storage, represents the largest difference with other constructs reflected by other span tasks since (1) processing is separated from storage in complex span tasks, and (2) processing is moderately involved in simple span tasks. One further hypothesis was that this compound variable should help characterize the nature of the relationship between WM capacity and general intelligence.


e1
e2
e3
e4
e5
OS
SS
SIM
MU
SSTM
.75
.54
.56
.66
.43
.86
.73
.75
.81
.66
Span Tasks
S and P
Combined
.73
-.22
.80
.44
Raven
Other

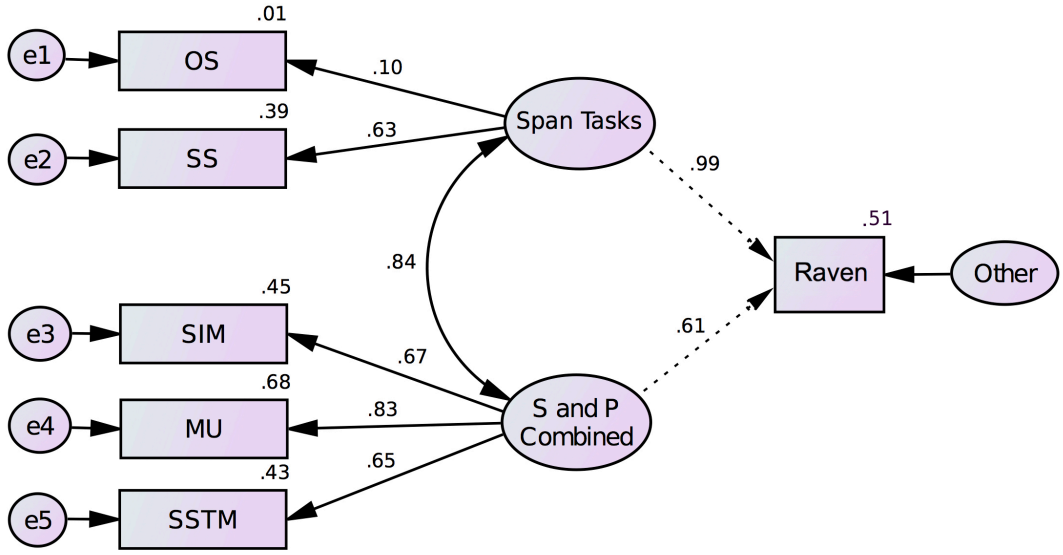
(a)


e1
e2
e3
e4
OS
SS
SIM
MU
.73
.55
.52
.67
.85
.74
.72
.82
Span Tasks
S and P
Combined
.75
-.26
.85
.46
Raven
Other

(b)


.74
e1
OS
.86
Span Tasks
.54
e2
SS
.73
.37
.44
.55
.60
Raven
Other
.55
e3
SIM
.46
.54
.50
.65
e4
MU
.54
S and P
Combined
.44
.48
e5
SSTM

(c)


.01
e1
OS
.10
Span Tasks
.39
e2
SS
.63
.99
.51
.84
Raven
Other
.45
e3
SIM
.61
.67
.68
e4
MU
.83
S and P
Combined
.65
.43
e5
SSTM

(d)

	$\chi 2$	CFI	RMSEA	RMR
A	$2.85(7), p = .9$	1	0	.002
B	$1.24(3), p = .75$	1	0	.020
C	$2.72(5), p = .74$	1	0	.020
D	$54.95(9), p = 0$	.735	.245	.216

(e)

Figure 16 – Path models from confirmatory factor analysis with (a) and without (b) the SSTM task. For comparison, a third model (c) used a factor reflecting all of the span tasks versus another factor reflecting only the three tasks in which *storage* and *processing* were combined. A fourth model (d) further constrained the parameters between Raven and the latent variables to 1 (dotted lines), and a table (e) recapitulates the fit of each model. Legend: OS = operation span; SS = sentence span; SIM = span task based on an adaptation of the SIMON® game; SSTM = spatial short-term memory; MU = memory updating; S and P, storage and processing; Raven = Raven's Advanced Progressive Matrices. The numbers above the arrows going from a latent variable (oval) to an observed variable (rectangle) represent the standardized estimates of the loadings of each task onto each construct. The paths connecting the latent variables (ovals) to each other represent their correlations. Error variances were not correlated. The numbers above the observed variables are the R^2 values.

Experiment 2

Experiment 2 was set up to better separate the *storage* and the *storage* × *processing* components by using complex versus simple sequences respectively. Two versions of our span task were created, one based on the most complex sequences of Experiment 1, and the second based on the simplest sequences of Experiment 1. The complex sequences targeted the *storage*-only demand because such sequences allow little or no compression. On the contrary, the simplest sequences targeted the *storage* × *processing* demand since the regularities could be used by participants to optimize storage.

Also, Experiment 1 estimated capacity from an inflexion point obtained across individuals who were not administered the same sequences of colors. Another goal of the second experiment was to provide estimates of the span based on a similar set of sequences given to the participants in progressive order (from sequences of minimal length to the longest length achieved) to ensure

a more proper evaluation of capacity. Performance on our span tasks was compared to other simple span tasks in a commercial test (WAIS-IV) and again to a measure of general intelligence (Raven's Advanced Progressive Matrices). Experiment 1 showed that complex span tasks were less correlated to a $storage \times processing$ construct, so we chose to focus on simple span tasks for which processing is partly dedicated to the storage process. It was hypothesized that the spans obtained from the Simon span tasks would better correlate with the Raven. Effectively, the spans of the WAIS either moderately involve processing (forward condition) or simply require linearly processing the stored items (backward or sequencing conditions). In the case of the backward and the sequencing conditions, processing cannot be regarded as subserving storage, because processing intervenes only after the items are stored (see Thomas, Milner, & Haberlandt, 2003, who showed that the sequences are initially stored before being processed). The span estimated during a chunking process is the only one that truly involves an optimization process based on the processing component being at the service of the storage component. If the span results from the participant's storage capacity which hierarchically depends on an optimization process (storage capacity is supposed to be fixed for a given individual, but any form of structuration of the stimulus sequence can contribute to extending the number of stored items), we hypothesize that our span tasks will best correlate with intelligence. The rationale is that during the Raven, participants also use storage and processing in conjunction to solve the problems.

Method

Participants

A total of 107 undergraduate students volunteered to participate in this experiment and received course credits in exchange for their participation ($M = 22.9$, $SD = 5.9$). All participants were administered our span task and three subtests measuring the span from the WAIS-IV; 95 of

the students agreed to be tested with the Raven's APM (which was optional for getting course credits). The average IQ for this sample was 100.2 ($SD = 13.4$).

Procedure

The tests were administered in the following order: the Simon span task, the WM subtests of the WAIS-IV and finally, and optionally, the Raven.

Simon span task

The procedure was designed to better match a standard memory span task in which a progression of difficulty is involved. Thus, the length of the presented sequences progressively increased, starting with length two, then three, etc. Two different sequences with a given length were presented (to match the number of repetitions per length in the WM subtests of the WAIS), and the session automatically stopped after the participant failed to correctly recall the two sequences of a given length. Scoring of the span followed that of the WAIS, in which the longest span attained at least once was considered as the subject's span.

Each participant was administered two complexity conditions (counterbalanced between participants). The sequences in each condition were taken from the 9150 sequences that were generated in Exp. 1. Based on the distribution of the sequences' given complexity, all the sequences were ranked according to their complexity (the 100th percentile corresponded to the most complex sequences), and a series of sequences was chosen at the 50th or the 100th percentile to constitute the two experimental conditions (the sequences below the 50th percentile were judged too easy to constitute an interesting experimental condition). By construction, the Simple condition was conducive to inducing a chunking process, while the Complex condition

allowed less chunking opportunities and as such was considered as mostly soliciting the storage component.

WAIS-IV working memory subtests

The participants were administered three WM subtests from the WAIS-IV (Wechsler, 2008): the Digit Span Forward (DSF) which requires recalling a series of digits in correct order, the Digit Span Backward (DSB) which requires recalling a series of digits in reverse order, and the Digit Span Sequencing (DSS) which requires recalling a series of digits in ascending order. The two last subtests clearly require mental manipulation of the stored items, in comparison to the first subtest which is considered the simpler short-term memory test. The scoring procedure followed the recommendation of the WAIS, the longest span attained at least once was considered the subject's span. Note that although these tasks are supposed to index a WM construct in the WAIS, these subtests are closer to simple span tasks and might better reflect a STM construct (see Unsworth & Engle, 2007b).

WAIS-IV working memory subtests

As in Exp. 1, the participants were asked to complete the matrices of set #2 which contained 36 matrices after doing set #1 which was used as a warm-up. The range of possible raw scores was 0-36.

Results

To estimate the reliability of our task, we split the data into two groups (odd vs. even trial numbers) within each sequence length. Participants showed equivalent performance at the even

Table 6 – *Stepwise forward selection model based on BIC criterion. At each stage (model), the variables are listed in increasing BIC order. The final model includes complexity and length, but no interaction term.*

Base model	Variable included	Deviance	BIC
Intercept only	Complexity	1672.9	1688.6
	Length	1737.3	1753.0
	None	2957.1	2965.0
Correct∼complexity	Length	1632.0	1655.6
	None	1672.9	1688.6
Correct∼Complexity & Length	None	1632	1655.6
	Interaction	1632	1663.5

and odd trials for each condition (Simple condition: $r = .62, p < .001$; Spearman-Brown coef. = .77; Complex condition: $r = .59, p < .001$; Spearman-Brown coef. = .74) and adequate evidence of reliability was obtained given such a short procedure. For comparison, we applied the same method to the digit span tasks, and we obtained estimates of reliability within the same range (Forward condition: $r = .45, p < .05$; Spearman-Brown coef. = .63; Backward condition: $r = .66, p < .001$; Spearman-Brown coef. = .78).

The mean complexity of the recalled sequences was again lower than that of the to-be-remembered sequences ($t(2653) = 13.20, p < .001$, Cohen's paired d = 0.26). Once again, we investigated the effect of length and complexity on perfect recall by way of logistic regression. A stepwise forward logistic model selection based on a BIC criterion suggested dropping the interaction term, as shown in Table 6.

A logistic regression on the 2654 trials was done on the proportion correct of perfectly recalled sequences as a function of complexity and length (see Figure 9). It confirmed the detrimental effect of complexity on recall ($z(2651) = 9.77, p < .001$, standardized coefficient = −6.41 [non standardized: −.95]) and also an effect of length ($z = 6.273, p < .001$, standardized coefficient = 3.94 [non standardized: 2.01948]). Overall this first result indicates that the memory process could effectively be optimized for less complex sequences. One interesting point is that the

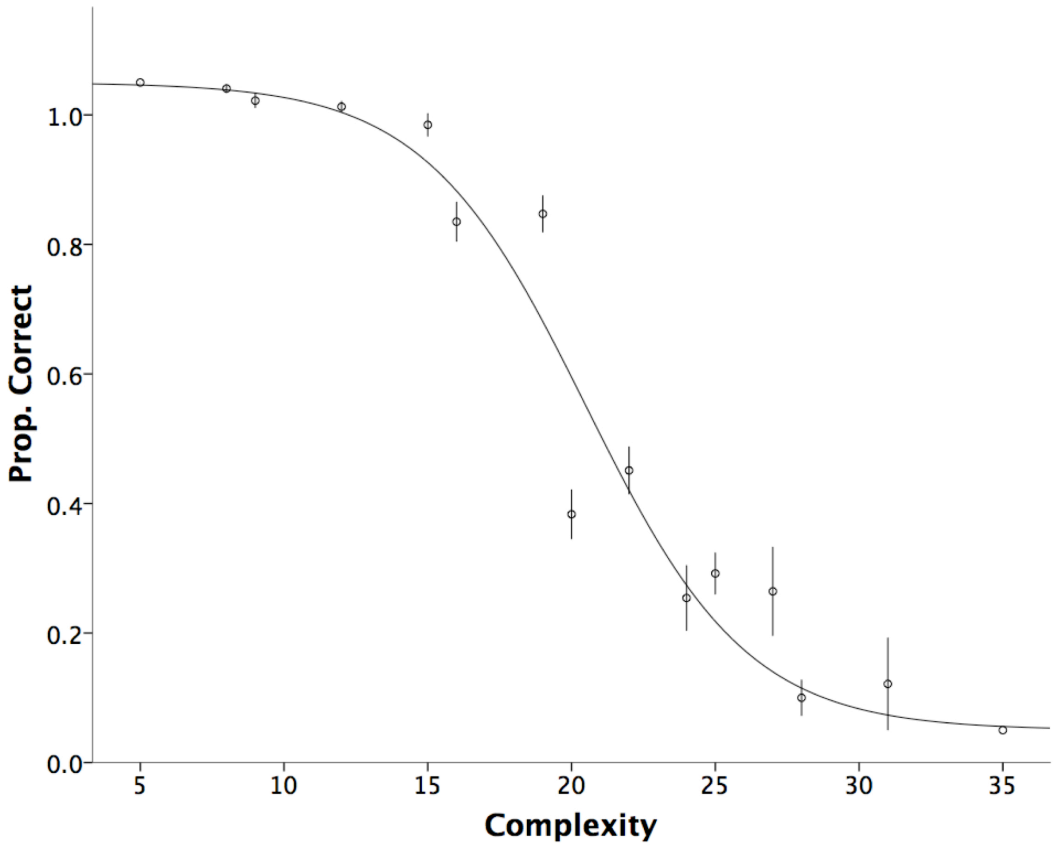


Figure 17 – Proportion correct as a function of complexity of the Simon span tasks (with the conditions Simple and Complex mixed) in Exp. 2. Error bars are +/- one standard error.

inflection curve is close to the one obtained in Experiment 1 (20.2 in comparison to 21.7 for Experiment 1), which means that allowing a greater amount of time to participants did not lead them to increase performance much. Still this difference was found statistically reliable ($t(270) = 4.5, p < .001$; Cohen's d = .50), although this difference must be tempered by the fact that the sequences of Exp. 2 were in general shorter (4.7 colors per sequence instead of 6.2 colors in Experiment 1, $t(5497) = 33, p < .001$; Cohen's d = .63). Also, the difference between the two inflection points only reflected a difference of .7 colors recalled (7 in Experiment 1 and 6.3 in Experiment 2).

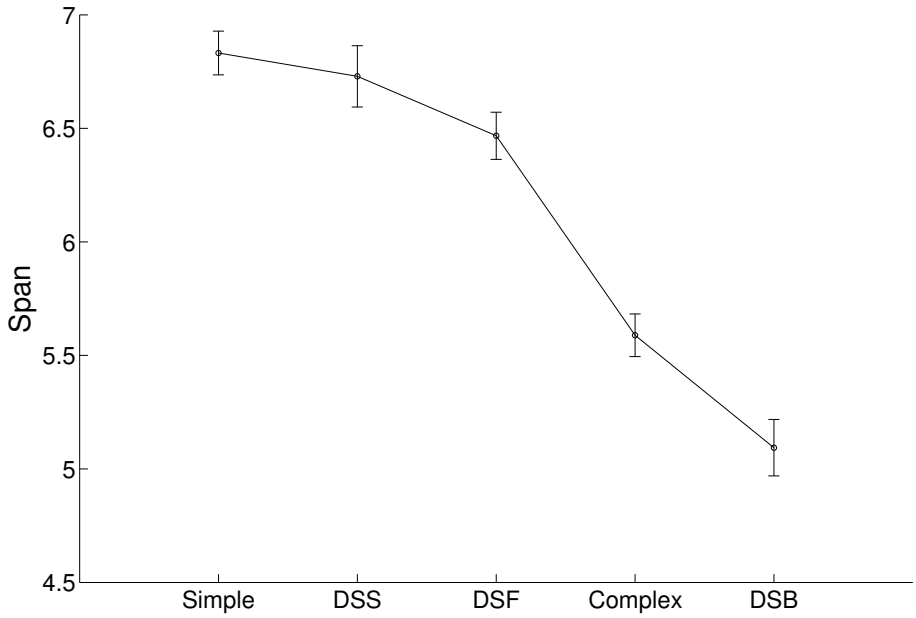


Figure 18 – Mean span as a function of type of span task in Exp. 2. Note: Simple, Simon span task, simple version; DSS, Digit Span Sequencing; DSF, Digit Span Forward; Complex, Simon span task, complex version; DSB, Digit Span Backward.

The decreasing mean spans reported in Figure 18 for DSB, Simon Complex, DSF, DSS, and Simon Simple are respectively 5.1 ($SD = 1.3$), 5.6 ($SD = 1.0$), 6.5 ($SD = 1.1$), 6.7 ($SD = 1.4$) and 6.8 ($SD = 1.0$). A repeated-measures ANOVA on the mean span was conducted with the type of span task as a within-subjects factor (Simon Simple, Simon Complex, DSF, DSB, DSS) on the data collapsed by participants. The result of the ANOVA was significant ($F(4, 424) = 68.7, p < .001, \eta_p^2 = .39$), indicating that almost 40% of the variance was accounted for by the type of span task factor.

Posthoc analysis (Bonferroni corrections were made for the pairwise comparisons) showed significant systematic pairwise differences except between Simon Simple and DSS, and between DSS and DSF. A noteworthy difference was found between the two respective mean spans for

Table 7 – *Note: Raven, Raven's Advanced Progressive Matrices raw scores; COMPL, mean span on the Simon span task in the complex condition (with chunking opportunities reduced); SIMPL, mean span on the Simon span task in the simple condition (with chunking opportunities induced); DSF, Digit Span Forward; DSB, Digit Span Backward; DSS, Digit Span Sequencing; ** $p < .001$*

	COMPL	DSF	DSB	DSS	Raven
SIMPL	.422**	.294**	.337**	.157	.413**
COMPL		.229*	.353**	.310**	.385**
DSF			.473**	.273**	.290**
DSB				.476**	.446**
DSS					.297**

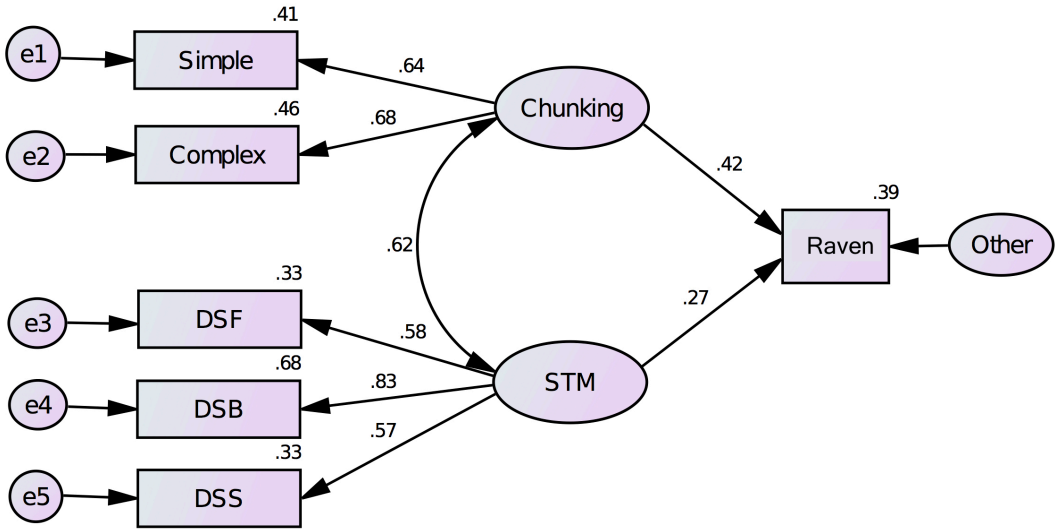
the two conditions in the Simon span tasks ($M = 1.25$; $SD = 1.1$), but surprisingly, the small difference of 1.25 on average means that between one and two additional colors were grouped on average within the simpler sequences.

Despite the moderate difference between the two mean spans observed between the simple and complex conditions, Table 7 shows that these two conditions highly correlated ($r = .42$), in comparison with other variables. Similarly, DSF and DSB shared the greatest percentage of variance ($r = .47$), as well as DSB and DSS ($r = .48$). Thus the digit spans showed high mutual correlation, but none of the digit span tasks correlated more with either Simon simple or Simon Complex than the two together. One possibility is that the participants could still chunk many of the most complex sequences, making the two Simon conditions akin, and accounting for the slight difference of 1.25 colors reported above. The possibility that participants chunk the less compressible sequences does not contradict compressibility theories because the estimate of the compressibility of a string is averaged across possible Turing machines. For a given complex string, a particular machine can still be able to perform compression. Another possibility is that participants verbalized the tasks in both conditions, and pronunciation of the sequences of colors was limited by the duration of rehearsal. In that case, chunking would not have been as beneficial as expected if only the number of colors to be pronounced determined capacity.

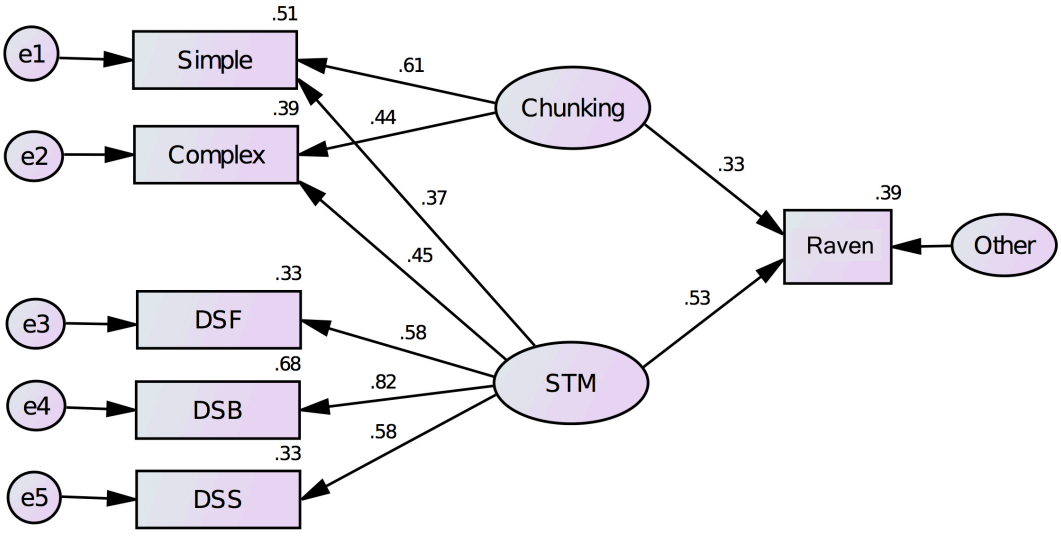
Regarding correlations with the Raven, the highest correlation was found with DSB, but the multicollinearity of the data makes interpretation of the pairwise correlations difficult.

The data were submitted to a confirmatory factor analysis using IBM SPSS AMOS 21 in order to test the prediction that tasks allowing the processing and storage components to fully function together in association to optimize storage are better predictors of general intelligence than the STM span tasks of the WAIS. A latent variable representing a chunking construct (derived from the Simon span tasks) and another latent variable representing a simpler STM construct (derived from the digit span tasks of the WAIS [14]) were sufficient to accommodate performance. The fit of the model shown in Figure 19 was excellent ($\chi 2(7) = 3.2, p = .87$; CFI, comparative fit index = 1.0; RMSEA, root mean squared of approximation = 0.0; RMR, root-mean square residual = .049; AIC and BIC criterions were both the lowest in comparison to a saturated model with all the variables correlated with one another and an independence model with all the variables uncorrelated), but again, as for Figure 16, with a caveat that the data fails to fit the recommended conditions for computing a CFA (because one factor is defined by only two indicators and because the correlation between the two factors reveal collinearity issues. These results suggests that the Raven is best predicted by the Chunking latent variable, a construct that can be reflected in the present study by the two span tasks. Figure 19b shows that a factor reflecting only the tasks for which Storage and Processing are supposed to be combined still account independently for the Raven once a first factor reflects all of the span tasks. The substraction of the less restrictive model (Figure 19c) from the more restrictive model (Figure 19a) showed that constraining the two paths to the Raven to be equal significantly reduced the model's ability to fit the data ($\chi 2(9 - 7 = 2) = 16.65 - 3.24 = 13.41$), meaning that the .42 loading in Model a can be considered significantly greater than $-.27$.

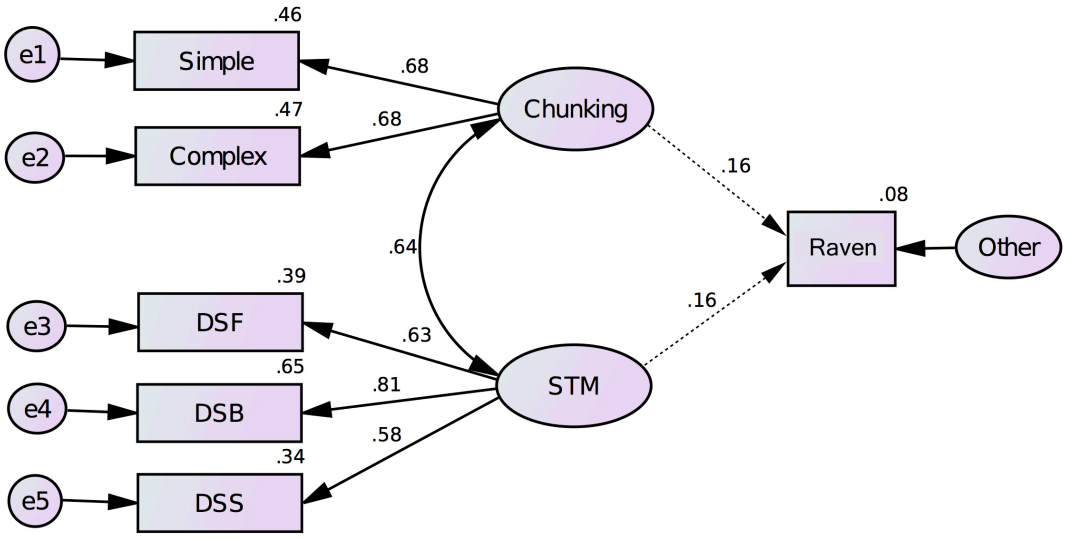
14. Again, we chose to label the latent variable STM instead of WM because the digit span tasks can be seen as simple span tasks (see Unsworth & Engle, 2007b).


e1
e2
e3
e4
e5
Simple
Complex
DSF
DSB
DSS
Chunking
STM
Raven
Other
.41
.46
.33
.68
.33
.64
.68
.58
.83
.57
.62
.42
.27
.39

(a)


e1
e2
e3
e4
e5
Simple
Complex
DSF
DSB
DSS
Chunking
STM
Raven
Other
.51
.39
.33
.68
.33
.61
.44
.37
.45
.58
.82
.58
.33
.53
.39

(b)



(c)

	$\chi 2$	CFI	RMSEA	RMR
A	$3.24(7), p = .86$	1	0	.003
B	$2.69(6), p = .85$	1	0	.003
C	$16.65(9), p = .05$	.93	.25	1

(d)

Figure 19 – Path models for confirmatory factor analysis from Exp. 2. An initial model (a) is compared with two other models: In (b), we used a factor reflecting all of the span tasks versus another factor reflecting only the three tasks in which storage and processing were combined in the task; In (c), we further constrained the parameters between the Raven and the latent variables to 1 (dotted lines). The table (d) recapitulates the fit of each of the models. All the correlations and loadings are standardized estimates. The numbers above the observed variables are the R-squared values. Simple, Simon Simple; Complex, Simon Complex; DSF, DSB and DSS, Digit Span Forward, Backward and Sequencing (WAIS-IV); Raven, Raven's Advanced Progressive Matrices; WM, traditional working memory construct; Chunking, chunking construct.

Discussion

The main goal of the second experiment was to provide several estimates of the STM span using similar scoring method. Performance on our span tasks was therefore compared to three other simple span tasks (Forward, Backward and Sequencing digit span tasks of the WAIS-IV) using the scoring method provided by the WAIS. The span task was hypothesized to best correlate with the Raven because both require to make sense of complexity while storing information, a process that might require optimizing storage capacity to free up memory and leave room for more complex reasoning. Experiment 2 was also devised to better separate the *storage* and the *storage* $\times$ *processing* demands by allowing two degrees of involvement of the processing component: low for the less compressible sequences that constituted the first condition (Complex version of the SIMON span task) and high for the second condition that included more compressible sequences (Simple version). The average span was found very close in the two conditions, with an average difference of only 1.2 additional colors recalled in the Simple condition of the span task than in the Complex condition. Although this difference was low, the correlation of the Simple version still slightly better correlated with the Raven than the Complex version. The verbalization of the colors could account for the fact that the span revolved around 7 for both the less compressible and the most compressible sequences, as if the participants were rehearsing the sequences of colors verbally without taking profit of the mathematical regularities (this is acknowledged as a limitation of our study in the General Discussion). Nevertheless, when the two tasks were used to estimate a latent variable covering a chunking process, the prediction of the Raven was found larger than for the three digit span tasks.

General discussion

Many complex mental activities still fit our rather low memory capacity. One observation is that complex concepts (those of experts or those transmitted by the culture) have been developed in the simplest possible way to become "simplex" (Berthoz, 2012). Pedagogy is also probably a quest for simplicity (for instance, the division algorithms first taught in universities are now taught in elementary school). But not all concepts are pre-compressed, and sometimes intelligence may require compression to achieve lower memory demands. We therefore tested a new conception of intelligence based on the idea that optimal behavior can be linked to compression of information. Our experimental setup was developed to study this compression process by allowing participants to mentally reorganize the to-be-remembered material in order to optimize storage in WM. To follow up on our previous conceptualization wherein WM is viewed as producing, on the fly, a maximally compressed representation of regular sequences (Chekaf et al., 2016 ; Mathy & Feldman, 2012), we used an algorithmic complexity metric to estimate the compressibility factor. Taken together, our results suggest that the existence of opportunities for compressing the memorandum enhance the recall process (for instance, when fewer colors are used to build a sequence, resulting in low complexity).

More interestingly, we found that having more repetitions of colors statistically interacted with sequence length, which indicates that the compressibility factor applies the best to longer sequences. This result recalls an earlier observation that longer sequences (i.e., greater than 5 or 6 items such as digits) are more likely to reveal a greater processing demand in simple span tasks (Unsworth & Engle, 2006, 2007b). Also, although length had a detrimental effect on recall, the complexity effect was found to be stronger than the length effect.

Regarding the relationship between compression and intelligence, the capacity to compress in-

formation (estimated by an individual's inflection point along the complexity axis of memory performance) was found to correlate with performance on Raven's matrices. This result indicates that participants who have a greater ability to compress regular sequences of colors tend to obtain higher scores on Raven's matrices. This correlation is comparable to the one obtained from a composite measure of WM capacity (using the WMCB). However, an exploratory analysis showed that our span task saturated two principal factors in a way very similar to that found for other tasks where processing is also devoted to storage (involving updating or free recoding of spatial information in the absence of a concurrent task), unlike the complex span tasks generally used to estimate fluid intelligence (remember that in complex span tasks, processing of the concurrent task is separate from storage in the main task, so it cannot support storage). A confirmatory factor analysis indicated greater predictability for the Raven test with a latent variable in which the storage and processing components were combined, in opposition to a latent variable representing complex span tasks in which storage and processing were separated. This latent variable shared 41% of the variance with Raven's matrices, which seems quite good compared to the 50% obtained by Kane et al. (2005), who used 14 different data sets (on more than 3000 participants) to estimate the variance shared between the WM-capacity and general-intelligence constructs (see also Ackerman, Beier, & Boyle, 2005, for a contrasting point of view). Our results confirm previous findings that memory updating can be a good predictor of Gf (e.g., Friedman et al., 2006 ; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009 ; Salthouse, 2014). More importantly, we think that the updating task could be associated with a SIMON span task under the same construct. Even though they do not seem to have much in common, processing (either updating or compressing) is dedicated to storage in both of these tasks.

These findings are largely consistent with previous studies (see Unsworth & Engle, 2007b),

suggesting that predicting fluid intelligence via simple spans with greater list lengths. Increasing digit list lengths, for instance, might require the participant to further process the sequence to optimize capacity. If the digit storage capacity is four independent digits (as in the brief visual presentations originally studied by Sperling, 1960), the only way to increase capacity is to recode or group together a few items.

The present study allows us to conclude that processing and storage should be examined together whenever processing is fully devoted to the stored items, and we believe that storage and processing must function together whenever there is information to be compressed in the memorandum. Thus, the ability to compress information in span tasks seems to be a good candidate for accounting for intelligence, along with other WM tasks (i.e., simple updating and short-term memory-span tasks) in which storage and processing also function together. This is in line with Unsworth et al. (2009), who argued that processing and storage should be examined together because WM is capable of processing and storing information simultaneously.

Cinquième partie

Conclusion Générale

Nous avons développé des tâches d'empan dans lesquelles stockage et traitement peuvent interagir pleinement. L'objectif était de mesurer une capacité en chunks de la mémoire de travail. La quantité de chunking a été estimée grâce à des mesures de complexité algorithmique de séries d'items à mémoriser. La maîtrise de la compressibilité (logique ou algorithmique) a permis un glissement de conceptions prenant en compte principalement le stockage vers la prise en compte d'une interaction *stockage* × *traitement*. D'après nos hypothèses de départ, davantage de possibilités de chunking impliquent davantage de place pour du traitement. Nous avons choisi de favoriser les processus de chunking dans le but d'éviter de séparer le traitement de sa matière, à savoir l'information à mémoriser. Ainsi, à partir de ces travaux il s'agit de déterminer si le chunking peut être un prédicteur fiable de l'intelligence lorsqu'il est mis à contribution dans des tâches d'empan simples. L'originalité de nos travaux est que nos tâches d'empan de chunks sont des tâches d'empan simples qui ciblent l'évaluation de la mémoire de travail, alors que traditionnellement, les tâches d'empan simples sont plutôt rattachées au concept de mémoire à court terme.

Tâches d'empan de chunks et intelligence fluide

Nous avons supposé que chunking et intelligence étaient tous deux liés aux capacités de stockage des individus. La capacité de stockage est censée être constante pour un individu donné, mais toute réorganisation d'une séquence à mémoriser peut contribuer à augmenter le nombre d'items stockés. Ce processus d'optimisation peut fonctionner par reconnaissance des régularités présentes au sein des matrices de Raven ou dans les séquences de couleurs ou d'objets à mémoriser. Pour résoudre les problèmes de la Raven ou chunker les items dans une tâche de mémorisation, les participants utilisent de façon conjointe leur capacité de stockage et de traitement. Au cours de plusieurs expériences, nous avons montré que l'empan de chunks était structurellement plus proche de la performance aux matrices de Raven que toutes les autres

tâches d'empan complexes de la working memory capacity battery (WMCB) ou qu'aucun des subtests de la WAIS. Parmi les tâches de la WMCB, la tâche de mise à jour de la mémoire et la tâche de mémoire à court terme spatiale sont celles qui ont montré le plus de variance partagée avec la tâche d'empan de chunks et avec la Raven, suggérant que la performance aux matrices de Raven est mieux prédite par une tâche d'empan lorsque traitement et stockage fonctionnent ensemble. Cela confirme un résultat de Mathy et Varré (2013) qui ont utilisé des listes de type alphanumériques et dans lesquelles les auteurs avaient introduit des répétitions d'items pour induire un processus de chunking. Leur tâche a montré de meilleures corrélations avec la tâche de mise à jour de la mémoire de la WMCB que des tâches similaires n'incluant pas de répétitions d'items. Un résultat majeur est que nos tâches d'empan de chunks rivalisent avec les tâches d'empan complexes dans lesquelles les items à mémoriser sont présentés de façon intercalée avec des tâches non liées à la mémorisation (voir Kane et al., 2005), comme par exemple des jugements de vérité sur des opérations arithmétiques simples ou sur des phrases. Les résultats de nos expériences montrent que des tâches d'empan simples peuvent rivaliser avec des tâches d'empan complexes en terme de corrélation avec des processus cognitifs de hauts niveaux. Ces résultats sont en cohérence avec ceux de Unsworth et Engle (2007b) qui suggèrent que lorsque la longueur des séquences dépasse 5 items, la prédiction de l'intelligence fluide par les tâches d'empan simples se rapproche de celle obtenue avec des tâches d'empan complexes. Haarmann, Davelaar, et Usher (2003) ont également montré une forte corrélation entre des tâches d'empan simples et l'intelligence fluide. Leur tâche d'empan de mots impliquait une réorganisation conceptuelle de l'information, les mots à mémoriser appartenaient à trois catégories sémantiques différentes. Haarmann et al. (2003) ont montré que l'empan conceptuel était mieux corrélé avec le raisonnement que l'empan de lecture. Selon eux, cette meilleure corrélation était liée au fait que cette tâche puisait dans des concepts de la mémoire à court terme sémantique. Kane et Miyake (2007) ont également souligné l'importance de la capacité

à faire des regroupements (clustering ability) pour les processus cognitifs de hauts niveaux. En effet, lorsque les mots à mémoriser étaient présentés par groupes, c'est à dire en clusters, la tâche ne prédisait plus l'intelligence. Par conséquent, il semble que ce soit le fait qu'une tâche d'empan pousse à la réorganisation de l'information qui en fait une tâche apte à prédire les hauts niveaux de cognition.

Mesurer traitement et stockage en interaction

Une conclusion majeure de ce travail de thèse est que traitement et stockage devraient être examinés de façon conjointe et ce, lors de situations où le traitement est entièrement dévolu aux éléments stockés. Nous croyons que c'est l'interaction entre stockage et traitement qui d'une part reflète le mieux le processus de chunking, et d'autre part fournit un véritable indice de la capacité globale de la mémoire de travail. Ainsi, parmi les différentes tâches d'empan, simples ou complexes, la tâche d'empan de chunks semble être un candidat précieux pour prédire l'intelligence et mesurer la mémoire de travail. Ceci est en cohérence avec les résultats de Unsworth et al. (2009) qui soutiennent que traitement et stockage devraient être examinés ensemble parce que la mémoire de travail est capable de traiter et de stocker des informations, simultanément. Towse, Cowan, Hitch, et Horton (2008) ont également mis en avant l'importance de l'interaction entre stockage et traitement, dans le cadre de l'hypothèse de reconstruction de rappel (recall reconstruction hypothesis). Cette hypothèse est basée sur des observations faites concernant la tâche d'empan de lecture (e.g., Cowan, Baddeley, Elliott, & Norris, 2003 ; Hitch, Towse, & Hutton, 2001). Dans cette tâche, les auteurs ont montré que les phrases (la partie traitement de la tâche) peuvent être utilisées pour reconstruire les mots cibles (la partie stockage de la tâche). Dans l'étude de Cowan et al. (2003), les auteurs ont observé que de longues pauses étaient souvent nécessaires pour pouvoir rappeler les mots. Towse et al. (2008) avancent deux affirmations importantes : 1) traitement et stockage ne doivent pas toujours

être considérés comme des événements complètement distincts, et 2) en mémoire de travail, traitement et stockage ne sont pas nécessairement des composantes qui sont en compétition pour les ressources cognitives (Barrouillet et al., 2004 ; Barrouillet, Plancher, Guida, & Camos, 2013 ; Case, 1985) ou qui interfèrent entre elles (Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012b ; Saito & Miyake, 2004). Selon l'hypothèse de reconstruction de rappel, traitement et stockage peuvent également interagir en synergie (e.g., Cowan et al., 2003 ; Hitch et al., 2001 ; Guida, Tardieu, & Nicolas, 2009). Lors d'une tâche d'empan de chiffres, Cowan et al. (2003) n'ont pas montré d'interaction entre traitement et stockage comme lors de l'empan de lecture. Nos tâches d'empan de chunks permettent d'étudier ce type d'interaction entre traitement et stockage en dehors du domaine linguistique, en condition d'empan simple.

Stratégies de compression de l'information

Il convient de noter que notre méthode vise à expliquer la formation immédiate de chunks à travers un processus de compression sans perte d'information, c'est à dire que les données compressées doivent permettre une récupération exacte des données originales (e.g., 'bleu-bleu-bleu' compressé en '3 bleu'). Les tâches sont appelées tâches d'empan de chunks en référence à une idée plus générale que des groupes perceptuels d'information peuvent potentiellement expliquer la formation de chunks conceptuels en mémoire à long terme, mais on ne s'attend pas à ce que tous les groupes d'information soient utiles pour la mémorisation en mémoire à long terme. Bien que le chunk '3 bleu' puisse être utile pour un essai, il ne le sera probablement plus à l'avenir. L'idée est que certaines formes de connaissance peuvent émerger au vol, au cours d'un processus de compression.

Mathy et Feldman (2012) ont montré que la capacité en mémoire de travail peut être évaluée à la fois en terme de chunks, dont le nombre est limité par la capacité stricte, et en terme

d'items contenus dans les chunks, le nombre d'items n'étant pas limité par la capacité. Par exemple, la séquence '1-2-3-4-2-2-2-2-6-5-4-3-2' peut être facilement divisée en trois séquences, en s'appuyant sur trois régularités présentes : '1234-2222-65432'. un participant qui rappelle les 13 items de cette séquence est largement au delà de la capacité de 7 ± 2 items, on peut supposer qu'il a profité de la compressibilité de la séquence, qui permet de compresser 13 items en trois chunks. Les tâches que nous développons visent à ne pas séparer le traitement du stockage. On ne contrôle donc pas les stratégies particulières des participants en les laissant libres de traiter l'information pour en optimiser le stockage. La présente thèse est basée sur l'hypothèse qu'un moyen de contrôler les stratégies individuelles est d'introduire des régularités systématiques dans le matériel à mémoriser en permettant aux participants de développer leurs propres stratégies. L'idée est que la compressibilité des séquences et les régularités introduites pour induire du chunking sont des moyens indirects de contrôle des stratégies particulières des participants, en effet, l'efficacité de ces stratégies doit refléter le niveau de compressibilité (de complexité) d'une séquence donnée.

Dans différentes études, nous avons montré que pour les séquences les moins compressibles, la capacité des participants était d'environ trois, c'est à dire la limite inférieure de la capacité de 4 ± 1 mesurée par Cowan (2001). Par ailleurs, la capacité des participants augmentait dès lors que des régularités ont été introduites dans les séquences à mémoriser, et donc lorsque la compressibilité indexée par différentes métriques de complexité algorithmique, augmentait. Dans toutes les expériences que nous avons conduites, l'empan des participants augmentait avec la compressibilité des séquences, ou lorsque l'ordre de présentation facilitait le recodage des séquences d'objets à mémoriser. Ces résultats sont le reflet de processus de recodage de l'information en mémoire de travail qui peuvent avoir lieu très rapidement. Les participants ont donc utilisé des stratégies de recodage de l'information et ces stratégies ont eu une efficacité

moyenne sur le rappel des séquences en lien avec les opportunités de compression. La longueur des séquences a également eu un effet (négatif) sur le rappel des participants, mais l'effet de la complexité sur le rappel était plus marqué que l'effet de la longueur (Chekaf, Gauvrit, Guida, & Mathy, sous presse, Exp.1 et 2).

Comme nous l'avons déjà évoqué, Feigenson et Halberda (2008) distinguent chunking conceptuel et perceptuel. Le chunking conceptuel consiste en un recodage utilisant des concepts connus, comme par exemple le fait de découper CIACNNBBC en CIA-CNN-BBC en se basant sur des connaissances en mémoire à long terme. Une autre forme de chunking est une forme de recodage basée sur les éléments à traiter. Feigenson et Halberda (2008) ont développé l'idée que la série "aubergine, tournevis, carotte, artichaut, marteau, pince" est plus facile à retenir que la série "aubergine, brocoli, carotte, artichaut, concombre, courgette". Les items de la deuxième série appartiennent pourtant tous à la catégorie des légumes, mais il semble plus aisé de retenir trois items de deux catégories distinctes plutôt que six items d'une même catégorie. Ici, le recodage se fait en référence à deux types conceptuels (légumes et outils) mais sans référence à des concepts pré-existants, comme par exemple un concept qui regrouperait l'ensemble "aubergine-carotte-artichaut". Dans ce cas, le processus de recodage est basé sur la sémantique, mais les auteurs ont développé une troisième définition de chunking qui fonde le processus de segmentation sur des informations spatio-temporelles (par exemple les numéros de téléphones, ou de cartes bancaires, regroupés en paires ou en quadruplets par proximité). Ce dernier type de chunking concorde avec les processus étudiés par Mathy et Feldman (2012), qui ont tenté de quantifier un processus de chunking immédiat en introduisant des patterns réguliers dans le matériel à mémoriser, sans lien avec des connaissances en mémoire à long terme.

Par ailleurs, dans les expériences que nous avons proposées, les couleurs ou les formes sont des concepts a priori déjà présents en mémoire à long terme pour les participants (au même titre

que carotte, tournevis et courgette), ce n'était pas l'enjeu de proposer des figures complexes inconnues comme l'ont fait Alvarez et Cavanagh (2004). En effet, les mesures algorithmiques de complexité que nous avons utilisées ne s'attachent pas du tout à caractériser les items individuellement, mais les relations entre les items. Ainsi, par exemple, jaune et rouge sont sans nul doute des concepts élémentaires partagés par tous les participants, mais la séquence 'jaune-rouge-jaune-rouge' peut facilement être recodée en une représentation plus compacte telle que '2 × jaune-rouge' dont la complexité dépend non pas des spécificités des items mais uniquement des liens entre les items.

Compression de données et mesure d'empan de chunks

La mesure d'empan de chunks teste la capacité des participants à traiter et stocker simultanément de l'information. L'idée est que les participants ayant des capacités de traitement plus élevées peuvent augmenter leur capacité de stockage en optimisant l'encodage, c'est à dire en adoptant des stratégies de chunking efficaces. Dans les tâches d'empan de chunks proposées, lorsque les séquences étaient compressibles ou présentaient des régularités, les participants ont effectivement montré des performances de rappel significativement plus élevées.

Les tâches mesurant l'empan de chunks évaluent directement la capacité de la mémoire de travail à stocker et à traiter l'information simultanément. Séparer les mesures du traitement et du stockage a pour effet de brouiller l'évaluation des capacités cognitives des participants, en empêchant une interaction capitale entre une capacité limitée de stockage et la manipulation de l'information dont un intérêt peut être d'optimiser l'espace de stockage. La composante de traitement est souvent négligée lors des tâches d'empan complexes dans lesquelles peu d'erreurs sont commises. La présente thèse visait à développer des tâches d'évaluation des capacités de la mémoire de travail laissant libres les participants de chunker le matériel à mémoriser, et

permettant, à condition de mesurer de manière fiable la complexité de ce matériel, d'évaluer des capacités cognitives de hauts niveaux. Ce type d'approche visait à une meilleure conceptualisation des liens entre les tâches d'empan simples mesurant la mémoire à court terme et les tâches d'empan complexes dont la vocation est d'évaluer la mémoire de travail. Nous pensons que ces tâches évaluent l'inverse de ce qui leur est théoriquement attribué. Les tâches simples évaluent selon nous davantage une interaction entre stockage et traitement, donc une mémoire immédiate libre de 'travailler', et focalisée sur la tâche de mémorisation. Les tâches d'empan complexes évaluent quant à elles une capacité brute de stockage (la composante de traitement est neutralisée, et rarement prise en compte dans les mesures d'empans complexes), en d'autres termes, un stockage dépourvu. Ces deux types de tâches ont montré des corrélations avec des évaluations d'intelligence fluide, de raisonnement, des aptitudes scolaires en lecture, écriture, mathématiques et sciences (Colom et al., 2004 ; St Clair-Thompson & Sykes, 2010), uniquement pour des séquences longues en ce qui concerne les tâches d'empan simples (Unsworth et al., 2009 ; Unsworth & Engle, 2007b). Dans les tâches d'empan de chunks proposées, le traitement peut être librement dirigé vers le matériel à mémoriser, avec l'idée que la composante de traitement agit plus efficacement lorsqu'elle est orientée vers l'information principale et peut compenser des capacités de stockage modestes par un encodage performant.

Complexité et compression de l'information en mémoire de travail

Dans cette thèse, la mesure de la compressibilité de l'information s'est appuyée sur des modèles théoriques de l'information (Shepard, Hovland, & Jenkins, 1961 ; Feldman, 2000 ; Delahaye & Zenil, 2012). La compressibilité des séquences d'objets à trois dimensions était mesurée par la complexité de Feldman, la compressibilité des séquences de couleurs était évaluée par une mesure de la complexité algorithmique de l'information adaptée aux listes de faible complexité (i.e., de longueur inférieure à 50 caractères). Ainsi, chacune des séquences à mémoriser présentait

un indice de complexité déterminé. La complexité correspond à la longueur de la formule la plus courte pour définir une catégorie. Considérons, par exemple, un concept booléen à trois dimensions, basé sur deux formes, deux tailles et deux couleurs possibles, s'il présente les trois exemples positifs suivants : ▲ ▲ △ (grand triangle noir, petit triangle noir, grand triangle blanc, autrement dit, les triangles s'ils sont grands ou noirs). La formule la plus courte pour définir l'ensemble des objets appartenant à cette liste est : triangles (grands ou noirs). Le nombre de traits réduit (au maximum) correspond à la complexité de l'ensemble (est ici égale à 3). Ce type de simplification est permis par une capacité à extraire les traits nécessaires et suffisants à la description d'un ensemble. Les performances de rappel étaient significativement améliorées lorsque le matériel était plus compressible, ce qui peut être expliqué par une réduction de la quantité d'information permise par la capacité des individus à l'abstraction.

Les mesures de complexité que nous avons utilisées se sont révélées adaptées aux séquences présentées, en effet, quelle que soit la longueur de la séquence, plus son indice de complexité augmentait, plus les performances de rappel diminuaient de manière significative. L'effet est plus marqué si l'on mesure les performances en fonction des indices de complexité de Feldman à longueur constante : si la complexité de Feldman augmente, alors la performance diminue significativement.

La complexité de Feldman a été un bon indicateur de la compressibilité du matériel présenté, à longueur constante mais aussi pour l'ensemble des séquences. La complexité permet d'évaluer le coût de l'encodage, plus la séquence a un indice de complexité élevé, plus l'encodage sera couteux. Un coût d'encodage élevé correspond à une moindre probabilité de chunker. Sur ce point, Brady et al. (2009) suggèrent que le chunking pourrait être astucieusement utilisé comme une approximation discrète d'un algorithme de compression idéale, c'est-à-dire qu'il pourrait être pensé comme mesure possible de la compressibilité psychologique. Ainsi, a pos-

teriori, à partir des données connues sur les capacités de la mémoire à court terme, le rappel moyen d'une séquence donnerait une indication de sa compressibilité et par extension, de sa complexité.

Régularités, ordre de présentation, chunking

Les manipulations du type d'ordre dans les séquences visaient à explorer les capacités de chunking. En effet, nous avons construit des conditions d'ordre en formulant l'hypothèse que certains favoriseraient la compression du matériel, et son rappel dans le bon ordre. L'effet des conditions d'ordre sur les performances de rappel, a montré, conformément à nos attentes, de meilleures performances lorsque les items étaient présentés en clusters distinguables (Chekaf et al., 2016). Ainsi, dans la condition règle, le rappel a été plus performant que dans la condition similarité, et encore plus que dans la condition dissimilarité. La similarité comme critère exclusif d'ordre entre objets successifs a conduit à de moins bonnes performances que lorsqu'on la combinait avec des critères de régularités. Dans la condition règle, des régularités étaient associées à des similarités dans l'ordre de présentation. Entre deux clusters, la similarité entre deux objets successivement présentés était faible. A l'intérieur d'un cluster la similarité était maximale avec une seule dimension de différence. De plus, les clusters d'une même séquence regroupaient des paires d'objets ayant une même différence et dans le même sens. Par exemple, une même différence de couleur à l'intérieur d'un cluster peut être perçue comme la règle 'blanc puis noir' si elle se retrouvait dans tous les autres clusters d'une même séquence. Ainsi, l'ordre de présentation : □ ■ □ ■ △ ▲ △ ▲ (grand carré blanc - grand carré noir, petit carré blanc - petit carré noir, petit triangle blanc - petit triangle noir) facilitait l'extraction de la dimension couleur en réduisant la séquence à [(grand carré, petit carré, petit triangle) blancs puis noirs]. Conformément à nos attentes, l'ordre basé sur la règle a conduit à de meilleurs résultats. En présentant les objets selon des régularités et dans des clusters distinguables, il a favorisé

la formation d'énoncés réduisant plus efficacement l'information présentée. Ce résultat est à comparer à celui obtenu par Mathy et Feldman (2009) pour une tâche d'apprentissage. L'ordre basé sur la règle était plus avantageux que celui basé uniquement sur la similarité entre stimuli. Dans la similarité, les objets étaient reliés entre eux sur des critères de proximité dimensionnelle pour former la séquence la plus fluide possible. Alors que les modèles de la mémoire à court terme fondés sur l'interférence prévoient que les performances des participants sont plus faibles lorsqu'ils perçoivent des similarités de traits dans les séquences à mémoriser, nos hypothèses allaient dans le sens de meilleures performances dans le cas de similarité entre objets que dans le cas de dissimilarité. Toutefois, si la similarité a conduit à de meilleures performances que la dissimilarité, elle a conduit à des performances moins élevées que pour l'ordre fondé sur des règles. Ces résultats suggèrent selon Mathy et Feldman (2009) que l'apprentissage humain implique un processus d'abstraction fondé sur des règles permettant un recodage algorithmique simple. L'ordre par règle correspond probablement mieux aux processus d'abstraction humains.

Limites et perspectives

Dans des tâches de rappel libre, il a été montré que la répétition d'items dans une même séquence conduit à une meilleure probabilité de leur rappel, avec un effet de la répétition amélioré par l'introduction d'espaces ou lorsque des items isolés les séparent (Kahana & Howard, 2005) comme s'il s'agissait de balises entre des espaces réguliers. Les participants cherchent à minimiser la distance du chemin qu'ils doivent parcourir à travers l'espace psychologique (Romney, Brewer, & Batchelder, 1993 ; Lewandowsky, Brown, & Thomas, 2009). Cette minimisation du chemin opère par un encodage sélectif focalisant sur les dimensions pertinentes entre les items similaires, ce qui pourrait être expliqué ici par une plus grande distinction temporelle des clusters. Le traitement étant plus rapide lorsque la charge informationnelle est réduite. Ainsi, l'encodage a pu être plus rapide lorsque les objets étaient associés en clusters, et plus long

entre deux objets appartenant à deux clusters différents. Une telle différence dans la rapidité du traitement pourrait être à l'origine d'une déformation de la représentation psychologique de la chronologie de succession des items. En résumé, une minimisation du chemin correspondant à une compression de l'information entraînerait des différences de vitesse de traitement selon les clusters ce qui favoriserait des distinctions temporelles. Ces dernières peuvent induire, par un mécanisme de déformation psychologique du temps, la formation de chunks.

L'introduction de pauses doit augmenter la distinguabilité temporelle et donc le chunking. De futures recherches pourraient proposer des variations temporelles dans la présentation, en association avec les types de structurations que nous avons proposés. Dans des séquences dans lesquelles des chunks sont induits par construction, par exemple en paires d'objets chunkables selon une règle simple, ajouter un facteur temporel, comme des pauses plus longues entre deux clusters, pour augmenter physiquement leur distinguabilité temporelle. Ou à l'inverse en introduisant des pauses à l'intérieur des clusters pour vérifier la robustesse d'une forme de chunking basé sur les regroupements logiques versus une forme de chunking basée sur les regroupements temporels. En effet, si les chunks peuvent être formés spontanément, certains facteurs, comme le regroupement spatial ou temporel de l'information, peuvent induire le chunking (Koch & Hoffmann, 2000 ; Lien & Ruthruff, 2004).

Par ailleurs, l'échelle de complexité algorithmique déjà utilisée récemment en psychologie (Gauvrit, Zenil, Delahaye, & Soler-Toscano, 2014 ; Gauvrit et al., 2015) est suffisamment fine et nous a déjà permis de mettre en place un protocole d'entrainement de la mémoire de travail, par paliers de complexité. Ce protocole est particulièrement intéressant puisqu'il s'adapte en temps réel aux performances des participants. Les séquences du test se basent sur le principe de séquences de couleurs à mémoriser. Une version séquentielle (les couleurs s'affichent successivement, les unes après les autres) et une version flash (toutes les couleurs d'une séquence

s'affichent en même temps sur l'écran) ont été mises au point. Les séquences les plus faciles sont de très faible complexité et de faible longueur, mais le test peut monter à des niveaux relativement élevés de complexité, ou de longueur de façon non dépendante, c'est à dire qu'au delà d'une longueur de quatre items, il est possible de faire varier la complexité de la tâche en maintenant la longueur constante, ou de maintenir un niveau de complexité en faisant varier la longueur de la séquence à mémoriser. Ce protocole pourra être décliné sous plusieurs formes, notamment dans le cadre de bilans ou de remédiations en neuropsychologie. En effet, ce test n'est pas conditionné par un nombre restreint de séquences particulières qui auraient été normées lors de sa mise en place. Ce test est fondé sur des niveaux de complexité et des longueurs de séquences. Ainsi, pour chaque niveau de complexité, il existe un nombre important de séquences (un même niveau de complexité pour une même longueur, par exemple à une longueur de 7 items peut correspondre à une quarantaine de séquences possibles). La probabilité qu'une même séquence soit présentée deux fois pour une même personne est réduite. Ainsi, en testant une capacité en terme de chunks, c'est à dire fondée à la fois sur la complexité et la longueur de séquence, ce test pourra être administré plusieurs fois à une même personne sans qu'un effet de test / re-test (d'apprentissage) n'influence les performances.

Conclusion

Les tâches que nous avons proposées ont permis une évaluation de la capacité de la mémoire de travail qui autorise et prend en compte le chunking. Nous avons montré que les associations basées sur des règles présentent un avantage sur celles uniquement basées sur des similarités. L'ordonnancement de l'information selon des règles simples et sa présentation en unités discriminables et compressibles, donc chunkables, facilitera davantage son traitement et sa mémorisation que si l'information est construite sur une présentation favorisant exclusivement les similarités. Dans ce travail de thèse, nous avons fait le lien entre des mesures de complexité

algorithmiques de l'information et différents types de mesures de la capacité de la mémoire immédiate. Nous avons relevé le paradoxe entre d'une part les conceptions théoriques de la mémoire de travail et de la mémoire à court terme, et d'autre part les outils de mesure de leurs capacités respectives. La mémoire à court terme est supposée représenter une forme de la mémoire immédiate restreinte à un stockage passif de l'information. Les tâches censées mesurer la capacité de la mémoire à court terme sont le plus souvent des tâches d'empan simples, dans lesquelles le rappel sériel libre est prédominant. C'est pourtant dans ce type de tâche que le matériel à mémoriser n'est pas directement contrôlé, et où les participants peuvent librement traiter l'information à mémoriser, sans que sa compressibilité ne soit connue. Ainsi, ces tâches ayant pour finalité de mesurer un stockage passif mesurent en réalité un stockage dans lequel un traitement tourné vers le stockage est permis. La mémoire de travail correspond quant à elle à une partie de la mémoire immédiate pour laquelle des processus de haut niveaux sont engagés afin de permettre au stockage et au traitement d'opérer de façon simultanée, le traitement pouvant être dirigé vers un autre matériel ou vers le matériel à mémoriser. Les tâches d'empan complexes sont des tâches classiques de la mesure des capacités de la mémoire de travail. Ces tâches ne sont pourtant pas construites de façon à mesurer une activité de traitement dédiée au stockage. Elles peuvent dans le meilleur des cas mesurer deux activités simultanées mais séparées, sans que l'activité de l'une ne bénéficie à l'autre.

Les tâches d'empan de chunks que nous avons proposées, grâce à l'apport considérable des mesures de complexité que nous avons eu l'opportunité de tester au cours de plusieurs expériences, ont montré leur efficacité à mesurer les capacités de la mémoire immédiate et permettent une synthèse entre mémoire à court terme et mémoire de travail. Lorsque la complexité des séquences est trop élevée, c'est à dire qu'elles sont peu ou non-compressibles, alors les individus n'ont pas l'opportunité de développer des stratégies simples et efficaces immédiatement pour

réduire l'information afin de tenter d'augmenter leur espace de stockage. Ce qui est mesuré dans ce cas, c'est une capacité de stockage imperméable au traitement, ce qui correspond à une capacité de stockage passif, ou une capacité brute de la mémoire à court terme, et qui se rapproche selon les estimations les plus optimistes de la capacité de 4 ± 1 chunks mesurée par Cowan (2001). Lorsque les séquences à mémoriser sont malléables et que leur complexité, si elle n'est pas trop élevée, permet aux individus d'adopter des stratégies de compression simples, et rapidement efficaces, alors elles mesurent la capacité de la mémoire de travail, c'est à dire la capacité $stockage \times traitement$.

Références Bibliographiques

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology*, *3*(301).

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological bulletin*, *131*(1), 30.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106-111.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341-380.

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728–748.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, *2*, 89-195.

Avons, S., & Mason, A. (1999). Effects of visual similarity on serial report and item recognition. *Quarterly Journal of Experimental Psychology: Section A*, *52*, 217-240.

Baddeley, A. (1986). *Working memory*. New York : Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2001). Is working memory still working? *American Psychologist*, *56*(11), 851.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation* (p. 647-667). New York, NY : Academic Press.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly journal of experimental psychology*, *36*(2), 233–252.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *14*, 575-589.

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83-100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570-585.

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science*, *21*, 413-419.

Barrouillet, P., Plancher, G., Guida, A., & Camos, V. (2013). Forgetting at short term: When do event-based interference and temporal factors have an effect? *Acta Psychologica*, *142*(2), 155–167.

Baum, E. B. (2004). *What is thought?* Cambridge, MA : MIT Press.

Berthoz, A. (2012). *Simplexity: Simplifying principles for a complex world (g. weiss, trans.).* Yale University Press.

Bor, D., Cumming, N., Scott, C. E., & Owen, A. M. (2004). Prefrontal cortical involvement in verbal encoding strategies. *European Journal of Neuroscience*, *19*(12), 3365–3370.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, *37*(2), 361–367.

Bor, D., & Owen, A. M. (2007). A common prefrontal-parietal network for mnemonic and mathematical recoding strategies within working memory. *Cerebral cortex*, *17*, 778-786.

Bor, D., & Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Frontiers in Psychology*, *3*(63).

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*, 201-233.

Boucher, V. J. (2006). On the function of stress rhythms in speech: evidence of a link with grouping effects on serial memory. *Language and speech*, *49*, 495-519.

Bradmetz, J., & Mathy, F. (2008). Response times seen as decompression times in Boolean concept use. *Psychological Research*, *72*, 211-234.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*, 487-502.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*, 1-34.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, *120*, 85-109.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.

Broadbent, D. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory.* New York : Wiley.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539-576.

Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551-581.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*(2), 229–246.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York, NY : Academic Press.

Chaitin, G. J. (1966). On the length of programs for computing finite binary sequences. *Journal of the ACM (JACM)*, *13*(4), 547–569.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, *33*, 1-16.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science. *Trends in Cognitive Sciences*, *7*(1), 19–22.

Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96 - 107.

Chekaf, M., Gauvrit, N., Guida, A., & Mathy, F. (sous presse). Compression in working memory and its relationship to fluid intelligence. *Cognitive Science*.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1235-1249.

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: the limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology*, *62*, 1420-1429.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295.

Colom, R., Jung, R. E., & Haier, R. J. (2007). General intelligence and memory span: Evidence for a common neuroanatomic framework. *Cognitive Neuropsychology*, *24*(8), 867–878.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: a reanalysis of key studies. *Memory & Cognition*, *34*, 158-171.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence*, *32*(3), 277–296.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory*, *14*(7), 804-813.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547-552.

Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing

adults. *Journal of Deaf Studies and Deaf Education*, *12*(3), 317–334.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-185.
Cowan, N. (2005). *Working memory capacity*. Hove, UK : Psychology Press.
Cowan, N. (2016). Working memory maturation can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*(2), 239-264.
Cowan, N., Baddeley, A., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: a comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review*, *10*, 74-79.
Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of experimental psychology. Learning, memory, and cognition*, *39*, 731-747.
Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to miller (1956). *Psychological Science*, *15*, 634-640.
Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100.
Cowan, N., Hismjatullina, A., AuBuchon, A. M., Saults, J. S., Horton, N., Leadbitter, K., & Towse, J. N. (2010). With development, list recall includes more chunks, not just larger ones. *Developmental Psychology*, *46*, 1119-1131.
Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*, 480-499.
Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology: Interdisciplinary and Applied*, *44*, 319-327.
Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the hebb effect. *Memory*, *11*(1), 43–63.
Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450 - 466.
Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.
Danziger, K. (2008). *Marking the mind: A history of memory*. Cambridge, UK: Cambridge University Press.
Davelaar, E. J. (2013). Short-term memory as a working memory control process. *Frontiers in Psychology*, *4*.
De Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Van Gorcum.
Delahaye, J.-P., & Zenil, H. (2012). Numerical evaluation of algorithmic complexity for short strings: A glance into the innermost structure of randomness. *Applied Mathematics and Computation*, *219*(1), 63–77.
Dieguez, S., Wagner-Egger, P., & Gauvrit, N. (2015). Nothing happens by accident, or does it? a low prior for randomness does not explain belief in conspiracy theories. *Psychological science*, *26*, 1762-1770.

Dixon, P., LeFevre, J.-A., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, *80*(4), 465.

Duncan, J. (2006, janvier). EPS mid-career award 2004: Brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology*, *59*(1), 2–27.

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review*, *19*(5), 864–870.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, *7*, 149-158.

Ehrlich, M.-F., & Delafoy, M. (1990). La mémoire de travail: Structure, fonctionnement, capacité. *L'Année Psychologique*, *90*(3), 403–427.

Elio, R., & Anderson, J. R. (1981). Effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 397-417.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, *12*, 20–30.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143-188.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19-23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309.

Ericsson, K. A., Chase, W., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning & Verbal Behavior*, *12*(4), 360-372.

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 128 - 145.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological review*, *119*, 223-271.

Farrell, S., Wise, V., & Lelièvre, A. (2011). Relations between timing, position, and grouping in short-term memory. *Memory & cognition*, *39*, 573-587.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 9926-9930.

Feldman, J. (1999). The role of objects in perceptual grouping. *Acta Psychologica*, *102*, 137-163.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630-633.

Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology*, *47*, 75-89.

Fernández, A., Ríos-Lago, M., Abásolo, D., Hornero, R., Álvarez-Linera, J., Paul, N., ... Ortiz, T. (2011). The correlation between white-matter microstructure and the

complexity of spontaneous brain activity: A difussion tensor imaging-MEG study. *Neuroimage*, *57*(4), 1300–1307.

Fernández, A., Zuluaga, P., Abásolo, D., Gómez, C., Serra, A., Méndez, M. A., & Hornero, R. (2012). Brain oscillatory complexity across the life span. *Clinical Neurophysiology*, *123*(11), 2154–2162.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*, 200-209.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*, 614-636.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychol Sci*, *17*, 172-179.

Gagné, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology*, *40*, 61-73.

Gauvrit, N., Singmann, H., Soler-Toscano, F., & Zenil, H. (2015). Algorithmic complexity for psychology: a user-friendly implementation of the coding theorem method. *Behavior research methods*, 1-16.

Gauvrit, N., Soler-Toscano, F., & Zenil, H. (2014). Natural scene statistics mediate the perception of image complexity. *Visual Cognition*, *22*(8), 1084–1091.

Gauvrit, N., Zenil, H., Delahaye, J. P., & Soler-Toscano, F. (2014). Algorithmic complexity for short binary strings applied to psychology: a primer. *Behavior research methods*, *46*, 732-744.

Gendle, M. H., & Ransom, M. R. (2009). Use of the electronic game simon® as a measure of working memory span in college age adults. *Journal of Behavioral and Neuroscience Research*, *4*, 1-7.

Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence*, *48*, 85 - 95.

Gilbert, A. C., Boucher, V. J., & Jemel, B. (2014). Perceptual chunking and its effect on memory in speech processing: Erp and behavioral evidence. *Frontiers in psychology*, *5*, 220.

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: recall of fewer but not smaller chunks in older adults. *Memory (Hove, England)*, *16*(7), 773-787.

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2009). Investigating the childhood development of working memory using sentences: new evidence for the growth of chunk capacity. *Journal of Experimental Child Psychology*, *104*, 252-265.

Gmeindl, L., Walsh, M., & Courtney, S. (2011). Binding serial order to representations in working memory: a spatial/verbal dissociation. *Memory & Cognition*, *39*, 37-46.

Gobet, F. (1998). Memory for the meaningless: How chunks help. In *Proceedings of the twentieth Annual Conference of the Cognitive Science Society* (p. 398-403). Mahwah, NJ : Erlbaum.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*,

5, 236-243.
Gobet, F., & Simon, H. A. (1996). Recall of random and distorted positions: Implications for the theory of expertise. *Memory and Cognition*, *24*, 493-503.
Goodwin, G. P., & Johnson-Laird, P. N. (2013). The acquisition of boolean concepts. *Trends in cognitive sciences*, *17*, 128-133.
Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322.
Guida, A., Gobet, F., & Nicolas, S. (2013). Functional cerebral reorganization: A signature of expertise? reexamining guida, gobet, tardieu, and nicolas'(2012) two-stage framework. *Frontiers in Human Neuroscience*, *7*.
Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. *Brain and cognition*, *79*, 221-244.
Guida, A., Tardieu, H., & Nicolas, S. (2009). The personalisation method applied to a working memory task: Evidence of long-term working memory effects. *European Journal of Cognitive Psychology*, *21*(6), 862–896.
Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: the roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*, 1001-1016.
Haarmann, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language*, *48*(2), 320–345.
Haladjian, H. H., & Mathy, F. (2015). A snapshot is all it takes to encode object locations into spatial memory. *Vision Research*, *107*, 133-145.
Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 325.
Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*, 80-115.
Hitch, G., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects and working memory: The role of the phonological loop. *Quarterly Journal of Experimental Psychology*, *49A*, 116-139.
Hitch, G., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*(2), 184.
Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *The Quarterly Journal of Experimental Psychology*, *25*, 104 - 111.
Hu, Y., & Ericsson, K. A. (2012). Memorization and recall of very long lists accounted for within the long-term working memory framework. *Cognitive Psychology*, *64*, 235-266.
Humes, L. E., & Floyd, S. S. (2005). Measures of working memory, sequence learning, and speech recognition in the elderly. *Journal of Speech, Language, and Hearing*

Research, *48*(1), 224–235.
Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organizatio and distinctiveness. *Journal of Memory and Language*, *32*(421-445).
Hutter, M. (2005). *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. Springer.
Jacobs, J. (1887). Experiments on "prehension". *Mind*(45), 75–79.
James, W. (1890). The methods and snares of psychology. In *The principles of psychology, vol i* (pp. 183–198). New York, NY, US : Henry Holt and Co.
Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science*, *20*, 568-577.
Johnson, N. F. (1969a). Chunking: Associative chaining versus coding. *Journal of Verbal Learning & Verbal Behavior*, *8*, 725-731.
Johnson, N. F. (1969b). The effect of a difficult word on the transitional error probabilities within a sequence. *Journal of Verbal Learning & Verbal Behavior*, *8*, 518-523.
Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, *144*, 1-13.
Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review*, *12*(1), 159-164.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on ackerman, beier, and boyle (2005). *Psychological Bulletin*, *131*(1), 66–71.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.
Kane, M. J., & Miyake, T. M. (2007). The validity of "conceptual span" as a measure of working memory capacity. *Memory & Cognition*, *35*(5), 1136–1150.
Karpicke, J. D., & Pisoni, D. B. (2000). Memory span and sequence learning using multimodal stimulus patterns: Preliminary findings in normal-hearing adults. *Research on Spoken Language Processing*.
Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span. *Memory & cognition*, *32*(6), 956-964.
Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247–254.
Kibbe, M. M., & Feigenson, L. (2014). Developmental origins of recoding and decoding in memory. *Cognitive Psychology*, *75*, 55-79.
Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, *112*(2), 240.
Koch, I., & Hoffmann, J. (2000). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological research*, *63*(1), 22-35.
Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, *1*, 1-7.

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, *37*(4), 347–364.

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology*, *51*, 57-74.

Lewandowsky, S., Brown, G., & Thomas, J. (2009). Traveling economically through memory space: Characterizing output order in memory for serial order. *Memory & Cognition*, *37*, 181-193.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review*, *11*, 771-790.

Lewandowsky, S., Oberauer, K., Yang, L. X., & Ecker, U. K. (2010). A working memory test battery for matlab. *Behavior Research Methods*, *42*, 571-585.

Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. New York, NY : Springler Verlag.

Lien, M.-C., & Ruthruff, E. (2004). Task switching in a hierarchical task structure: evidence for the fragility of the task repetition benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 697.

Lin, P. H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Vis cogn*, *17*, 356-372.

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in psychology*, *5*.

Logan, G. D. (2004). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General*, *133*, 218-236.

Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition*, *31*, 790-799.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.

Machado, B. S., Miranda, T. B., Morya, E., Amaro Jr, E., & Sameshima, K. (2010). Algorithmic complexity measure of EEG for staging brain state. *Clinical Neurophysiology*, *121*, S249–S250.

Majerus, S., Perez, M., & Oberauer, K. (2012). Distinct long-term memory mechanisms contribute to verbal short-term memory. *Journal of Memory and Language*, *66*, 38-51.

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple'short-term storage? *Intelligence*, *39*(6), 473–480.

Mathy, F. (2010). The long term effect of relational information in type VI concepts. *European Journal of Cognitive Psychology*, *22*, 360-390.

Mathy, F., & Bradmetz, J. (2004). A theory of the graceful complexification of concepts and their learnability. *Current Psychology of Cognition*, *22*, 41-82.

Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, *16*, 1050-1057.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*, 346-362.

Mathy, F., & Varré, J. S. (2013). Retention-error patterns in complex alphanumeric serial-recall tasks. *Memory*, *21*, 945-968.

Maybery, M. T., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language*, *47*(3), 360-385.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, *13*, 307-325.

Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, *1*, 250-254.

Melcher, J. M., & Schooler, J. W. (2004, Jun). Perceptual and conceptual training mediate the verbal overshadowing effect in an unfamiliar domain. *Memory & cognition*, *32*(4), 618–631.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology*, *56*, 485-491.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co.

Moher, M., Tuerk, A. S., & Feigenson, L. (2012). Seven-month-old infants chunk items in memory. *Journal of Experimental Child Psychology*, *112*(4), 361–377.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251-269.

Nairne, J. S. (2002). Remembering over the short-term: the case against the standard model. *Annual Review of Psychology*, *53*, 53-81.

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, *27*, 45-53.

Naveh-Benjamin, M., Cowan, N., Kilb, A., & Chen, Z. (2007). Age-related differences in immediate serial recall: Dissociating chunk formation and capacity. *Memory & Cognition*, *35*, 724-737.

Newell, A., & Simon, H. A. (1956). The logic theory machine–a complex information processing system. *IRE Transactions on Information Theory*, *2*(3), 61-79.

Ng, H. L., & Maybery, M. T. (2002). Grouping in short-term verbal memory: is position coded temporally? *Quarterly Journal of Experimental Psychology*, *55*, 391-424.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, *17*(9), 781–787.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411.

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & cognition*, *41*, 1212-1227.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*, 601-626.

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language*, *51*(1), 80–96.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012a). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012b). Modeling working memory: an interference model of complex span. *Psychonomic Bulletin & Review*, *19*, 779-819.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Perlman, A., Pothos, E. M., Edwards, D. J., & Tzelgov, J. (2010). Task-relevant chunking in sequence learning. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 649-661.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *Journal of Experimental Psychology: General*, *119*, 264-275.

Plummer, P., & Eskes, G. (2015). Measuring treatment effects on dual-task performance: a framework for research and clinical practice. *Frontiers in Human Neuroscience*, *9*, 225.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, *21*, 156-161.

Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Frontiers in Psychology*, *3*(113).

Rabinovich, M. I., Varona, P., Tristan, I., & Afraimovich, V. S. (2014). Chunking dynamics: Heteroclinics in mind. *Frontiers in Computational Neuroscience*, *8*(22).

Raven, J. C. (1962). *Advanced progressive matrices: Sets i and II.* HK Lewis.

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic bulletin & review*, *23*, 271–277.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*, 465–471.

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). Mdlchunker: A mdl-based cognitive model of inductive learning. *Cognitive Science*, *35*, 1352-1389.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, *4*, 28-34.

Ryabko, B., Reznikova, Z., Druzyaka, A., & Panteleeva, S. (2013). Using ideas of kolmogorov complexity for studying biological texts. *Theory of Computing Systems*, *52*(1), 133–147.

Ryan, J. (1969). Grouping and short-term memory: different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*, 137-147.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*(4), 425–443.

Salthouse, T. A. (2014). Relations between running memory and fluid intelligence. *Intelligence*, *43*, 1–7.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: the gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1089.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, 13, whole No. 517.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3-22.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*, 1244-1286.

Smyth, M. M., Hay, D. C., Hitch, G., & Horton, N. J. (2005). Serial position memory in the visual–spatial domain: Reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology: Section A*, *58*, 909-930.

Soler-Toscano, F., Zenil, H., Delahaye, J. P., & Gauvrit, N. (2013). Correspondence and independence of numerical evaluations of algorithmic information measures. *Computability*, *2*(2), 125–140.

Soler-Toscano, F., Zenil, H., Delahaye, J. P., & Gauvrit, N. (2014). Calculating kolmogorov complexity from the output frequency distributions of small turing machines. *PloS One*, *9*(5), e96223.

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., & Botvinick, M. M. (2014). Optimal behavioral hierarchy. *PLoS Computational Biology*, *10*, 1-10.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*(1-29).

St Clair-Thompson, H., & Sykes, S. (2010). Scoring methods and the predictive ability of working memory tasks. *Behavior Research Methods*, *42*, 969-975.

Szmalec, A., Page, M., & Duyck, W. (2012). The development of long-term lexical representations through hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342–354.

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall different response time patterns, same retrieval order. *Psychological Science*, *14*(2), 169–174.

Towse, J. N., Cowan, N., Hitch, G., & Horton, N. J. (2008). The recall of information from working memory: Insights from behavioral and chronometric perspectives. *Experimental Psychology*, *55*(6), 371.

Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Experimental Psychology*, *16*, 83-95.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*, 68-80.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*, 1038-1066.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: a latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635-654.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological science*, *21*, 384-390.

Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology*, *50*, 501-510.

Wang, B., Cao, X., Theeuwes, J., Olivers, C. N., & Wang, Z. (2016). Separate capacities for storing different features in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Wang, T., Ren, X., Li, X., & Schweizer, K. (2015). The modeling of temporary storage and its effect on fluid intelligence: Evidence from both brown–peterson and complex span tasks. *Intelligence*, *49*, 84–93.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*(2), 89.

Wechsler, D. (2003). Wechsler intelligence scale for children–fourth edition (wisc-iv). *San Antonio, TX: The Psychological Corporation*.

Wechsler, D. (2008). Wechsler adult intelligence scale–fourth edition (WAIS–IV). *San Antonio, TX: NCS Pearson*.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48-64.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*, 913-934.

Yagil, G. (2009). The structural complexity of DNA templates—implications on cellular complexity. *Journal of Theoretical Biology*, *259*(3), 621–627.

Zhang, G., & Simon, H. A. (1985). Stm capacity for chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, *13*, 193-201.

Liste des Figures

1 (a) : Dans ce diagramme de Hasse, chaque sommet correspond à un objet, chaque objet est défini par trois dimensions, forme, taille, couleur. Pour l'exemple, nous avons choisi les formes (carré et triangle), les couleurs (noir et blanc) et les positions sur le diagramme (grands en haut, noirs à droite, triangles au premier plan). Cependant, pour les expériences développées plus loin, les formes, les couleurs et les positions sont tirées au sort. (b) et (c) : exemple de deux concepts, en se référant à l'exemple (a), les objets pointés en (b) correspondent à "triangles et grands" et les objets pointés en (c) correspondent à "petits ou carrés". 31

2 Screen shots showing a sample of possible stimuli (top panel), a to-be-recalled patterned sequence of stimuli (middle panel), and an example of a response screen that asked the participant to recognize the previous stimuli and to order them by clicking on a mouse (bottom panel). 49

3 Example of three sequences from the category structure "square or small", representing six items (top left cube). *Note.* The six items are replaced by black numbered circles on each of the three cubes to simulate their sequencing. In each of the three cubes, the presentation order is indicated by the numbers from one to six and by the arrows. The distance between two consecutive objects is described by the type of arrow: solid (one edge, or one feature difference), dashed (two edges), or dotted (three edges). In the Rule condition, the objects are presented in three clusters, within which the solid arrows are parallel and go in the same direction. The regularity of the rule is related to two factors: the separation of the clusters (dashed arrows) and the similarity between the objects in the same cluster (solid arrows), both of which facilitate the formation of small groups. In the Similarity condition, the inter-stimulus distance is minimal. All the objects are linked to each other by solid arrows, which can potentially make a unique chunk of six objects. In the Dissimilarity condition, the sequence is characterized by a maximal inter-stimulus distance. The three plots at the bottom show the distances between the stimuli as a function of presentation time (1 second per item). For instance, the first plot shows two distance leaps in the Rule condition, which can facilitate the identification of three chunks. The leaps are more numerous, larger, and less regular in the Dissimilarity condition, while there are no leaps at all in the Similarity condition. 52

4 Proportion of correct sequences recalled as a function of the sequence length (number of objects), for each presentation order. *Note.* The error bars show ±1 standard error. 57

5 Proportion of correct sequences recalled as a function of Feldman's Complexity (FC), across each sequence length (left plot) and collapsed (right plot). *Note.* The analysis shown in the right plot is based on the 3217 trials, to include the entire variance. The error bars show ± one standard error. 60

6 Feldman's (2000) minimization of Boolean complexity. *Note.* Column FC indicates Feldman's (2000) measure of logical compressibility for each of the categories in the left or right column. 75

7 (Top) A sample of stimuli based on eight shapes, eight colors and two types of sizes. (Bottom) Table showing a sample of trials in Exp. 1 and in Exp. 2. The first half of the table shows the sequences of the non-chunkable condition used in both Exp. 1 and in Exp. 2. The second half of the table shows the sequences of the chunkable condition used in Exp. 1 exclusively. For each sequence length, we chose to represent only two trials. Dimension values were chosen randomly for each trial. For instance, the given category structure □ ■ □■ (i.e., square, small first, white-black within sizes) could be converted into either △ □ △□ (i.e., white, small first, triangle-square within sizes), or ■▲□△ (i.e., large first, square-triangle within colors), etc. The preceding example only involve the dimension values square, triangle white and black, but again, the dimension values were actually randomly picked among the 8 shapes and the 8 colors shown in the top panel. 87

8 Mean span for Experiment 1. 96

9 Histograms of the the mean spans observed in Experiment 2A (top left), and for the second phase of Experiment 2B (top right), for the second phase of Experiment 2C (bottom left), and in Experiment 2D (bottom right). *Note.* The spans were computed using the all-or-nothing-unit-scoring method (i.e., one point was cumulated for each perfectly recalled sequence, and the total was divided by the number of repetition per list length. 98

10 Stimuli used by Alvarez and Cavanagh (2004), and in Exp. 2C. From top to bottom: colors, alphabet, objects, kanjis, polygons, cubes. 106

11 Mean spans for Experiment 2C. Note: the results plotted in the boolean condition represent the span measured during the first phase of the task, which involved no specific instruction. 107

12 Example of a sequence of three colors for the memory span task adapted from the SIMON® game. 123

13 Proportion of perfectly recalled color sequences, by sequence length and number of different colors within a sequence. Note: Error bars show ± one standard error. 126

14 Proportion of perfectly recalled sequences of colors as a function of complexity. Note: Error bars shows ± one standard error. 129

15 Proportion of correct recall as a function of complexity, by sequence length. Error bars are ± one standard error. 130

16 Path models from confirmatory factor analysis with (a) and without (b) the SSTM task. For comparison, a third model (c) used a factor reflecting all of the span tasks versus another factor reflecting only the three tasks in which *storage* and *processing* were combined. A fourth model (d) further constrained the parameters between Raven and the latent variables to 1 (dotted lines), and a table (e) recapitulates the fit of each model. Legend: OS = operation span; SS = sentence span; SIM = span task based on an adaptation of the SIMON® game; SSTM = spatial short-term memory; MU = memory updating; S and P, storage and processing; Raven = Raven's Advanced Progressive Matrices. The numbers above the arrows going from a latent variable (oval) to an observed variable (rectangle) represent the standardized estimates of the loadings of each task onto each construct. The paths connecting the latent variables (ovals) to each other represent their correlations. Error variances were not correlated. The numbers above the observed variables are the R^2 values. 135

17 Proportion correct as a function of complexity of the Simon span tasks (with the conditions Simple and Complex mixed) in Exp. 2. Error bars are +/- one standard error. 140

18 Mean span as a function of type of span task in Exp. 2. Note: Simple, Simon span task, simple version; DSS, Digit Span Sequencing; DSF, Digit Span Forward; Complex, Simon span task, complex version; DSB, Digit Span Backward. . . . 141

19 Path models for confirmatory factor analysis from Exp. 2. An initial model (a) is compared with two other models: In (b), we used a factor reflecting all of the span tasks versus another factor reflecting only the three tasks in which storage and processing were combined in the task; In (c), we further constrained the parameters between the Raven and the latent variables to 1 (dotted lines). The table (d) recapitulates the fit of each of the models. All the correlations and loadings are standardized estimates. The numbers above the observed variables are the R-squared values. Simple, Simon Simple; Complex, Simon Complex; DSF, DSB and DSS, Digit Span Forward, Backward and Sequencing (WAIS-IV); Raven, Raven's Advanced Progressive Matrices; WM, traditional working memory construct; Chunking, chunking construct. 145

Liste des Tableaux

1 *Average proportion (all participants pooled) of adjacent two-stimulus sequences recalled, as a function of presentation order and pair type.* 63

2 *Mean span (and standard errors) in Experiment 1, by procedure (simple vs. complex span tasks) and category set complexity (chunkable vs. non-chunkable).* 95

3 *Mean Span for nonchunkable material and simple span task procedure, measured in Experiment 1, 2A, 2B, 2C, and 2D for adult participants, compared to an average span of 3 using one-sample t-tests. For Exp. 2A and 2D, the average span of the unique condition is reported. For Exp. 2B and 2C, the highest span measured across all conditions is reported to run a conservative test.* 99

4 *Mean Spans (and std errors) measured in Exp. 2C by Age-group by phase and by instruction type (verbalize vs visualize), for the boolean stimuli.* 105

5 *Stepwise forward selection of several models based on criterion BIC. At each stage, for a given model (where model $\sim v$ means that the base model is computed as a function of v to predict correct performance), the variables are listed in increasing BIC order after other variables were introduced. The final model includes complexity and length, but no interaction term. For instance, in the last model computed as a function of complexity and length, adding the interaction term increased the BIC, so the simplest model was chosen.* 128

6 *Stepwise forward selection model based on BIC criterion. At each stage (model), the variables are listed in increasing BIC order. The final model includes complexity and length, but no interaction term.* 139

7 *Note: Raven, Raven's Advanced Progressive Matrices raw scores; COMPL, mean span on the Simon span task in the complex condition (with chunking opportunities reduced); SIMPL, mean span on the Simon span task in the simple condition (with chunking opportunities induced); DSF, Digit Span Forward; DSB, Digit Span Backward; DSS, Digit Span Sequencing; ** $p < .001$* 142

Résumé

Les tâches d'empan simples sont classiquement utilisées pour évaluer la mémoire à court terme, tandis que les tâches d'empan complexes sont généralement considérées comme spécifiques de la mémoire de travail. Par construction, les tâches d'empan complexes étant plus difficiles, en raison de l'utilisation d'une tâche concurrente, l'empan moyen est généralement plus faible (4 ± 1 items) que dans les tâches d'empan simples (7 ± 2 items). Une raison possible de cet écart est que sans tâche concurrente, les participants peuvent tirer profit du temps libre entre les stimuli pour détecter et recoder des régularités présentes dans la série de stimuli afin de regrouper les stimuli en 4 ± 1 chunks. Notre hypothèse principale est que la compression de l'information en mémoire immédiate est un indicateur pertinent pour étudier la relation entre la capacité de mémoire immédiate et l'intelligence fluide. L'idée est que les deux dépendent de la qualité du traitement de l'information, autrement dit, selon nos hypothèses, d'une interaction entre traitement et stockage. Nous avons développé plusieurs tâches de mesure d'empan de chunks dans lesquelles la compressibilité de l'information a été estimée en utilisant différentes mesures de complexité algorithmique. Les résultats ont montré que la compressibilité peut être utilisée pour prédire la performance en mémoire de travail, et que la capacité à compresser l'information est un bon prédicteur de l'intelligence fluide. Nous concluons que la capacité à compresser l'information en mémoire de travail est la raison pour laquelle à la fois traitement et stockage de l'information sont liés à l'intelligence.

Mots clés : Chunking, Mémoire à Court Terme, Mémoire de Travail, Compression de l'Information, Intelligence.

Abstract

Simple span tasks are tasks commonly used to measure short-term memory, while complex span tasks are usually considered typical measures of working memory. Because complex span tasks were designed to create a concurrent task, the average span is usually lower (4 ± 1 items) than in simple span tasks (7 ± 2 items). One possible reason for measuring higher spans during simple span tasks is that participants can take profit of the spare time between the stimuli to detect, and recode regularities in the stimulus series (in the absence of a concurrent task), and such regularities can be used to pack a few stimuli into 4 ± 1 chunks. Our main hypothesis was that information compression in immediate memory is an excellent indicator for studying the relationship between immediate-memory capacity and fluid intelligence. The idea is that both depend on the efficiency of information processing, and more precisely, on the interaction between storage and processing. We developed various span tasks measuring a chunking capacity, in which compressibility of memoranda was estimated using different algorithmic complexity metrics. The results showed that compressibility can be used to predict working-memory performance, and that fluid intelligence is well predicted by the ability to compress information. We conclude that the ability to compress information in working memory is the reason why both manipulation and retention of information are linked to intelligence.

Keywords : Chunking, Short Term Memory, Working Memory, Information Compression, Intelligence.