Université de Bourgogne
Laboratoire d'Etude de l'Apprentissage et du Développement
C.N.R.S. – U.M.R. 5022

Ecole doctorale Environnements – Santé (ES)

THÈSE
Pour obtenir le grade de
Docteur de l'Université de Bourgogne
Discipline : **Psychologie**

L'émergence et l'évolution du caractère obligatoire des automatismes cognitifs

Présentée par **Laurent GREGOIRE**

Sous la direction de **Pierre PERRUCHET**
Co-dirigée par **Bénédicte POULIN-CHARRONNAT**

Thèse soutenue publiquement le 13 décembre 2013, devant le jury composé de :
Mireille BESSON, DR CNRS, Université d'Aix-Marseille, Présidente
Patrick BONIN, Professeur, Université de Bourgogne, Examinateur
Ludovic FERRAND, DR CNRS, Université de Clermont-Ferrand 2, Rapporteur
Olivier KOENIG, Professeur, Université de Lyon 2, Rapporteur

Remerciements

J'adresse en premier lieu mes plus sincères remerciements à Pierre Perruchet pour m'avoir offert l'opportunité de réaliser cette thèse. J'ai ainsi eu le privilège de profiter de sa remarquable disposition à conduire des recherches scientifiques.

J'exprime également toute ma gratitude à Bénédicte Poulin-Charronnat pour avoir initié et co-encadré ce projet. Je mesure la chance qui m'a été donnée de pouvoir compter sur son implication, sa disponibilité, et sa bienveillance tout au long de ce travail. Sa confiance et ses encouragements ont représenté, sans aucun doute, mon soutien le plus précieux.

Je tiens à remercier Mireille Besson, Patrick Bonin, Ludovic Ferrand et Olivier Koenig pour m'avoir fait l'honneur d'être membres de mon jury et d'estimer mon travail.

Je souhaite faire part de ma reconnaissance aux responsables des établissements et aux enseignants qui m'ont ouvert leurs portes, pour leur accueil, leur organisation, et leur intérêt envers mes recherches. Plus particulièrement, à Arnaud Péruta, directeur du conservatoire à rayonnement régional de Dijon, et Marc-Erec Saint-Jean-Vitus, directeur de l'école de musique de Longvic, ainsi qu'à Sacha Duchaine, Christelle Rithié, et Claudine Venitucci.

Et un grand merci à tous les enfants qui ont participé avec enthousiasme et application à mes expériences.

J'ai eu le plaisir de côtoyer de nombreux doctorants, ingénieurs et chercheurs durant mes années de thèse, dans le cadre de mes recherches ou des enseignements dispensés à l'université de Bourgogne. Je voudrais leur témoigner toute ma sympathie pour les échanges, parfois prolifiques, que nous avons entretenus, et surtout pour leur gentillesse. Je pense notamment à Jean-Yves Baudouin, Aurélia Bugaiska, Laura Ferreri, Cédric Jeanguyot, Yannick Glady, Aurélie Lagarrigue, Lolita Mercadié, Mélanie Provost, et Arnaud Witt.

Je tiens enfin à adresser mes remerciements aux organismes qui ont financé ce projet : l'agence nationale de la recherche (09-JCJC-0129) et le conseil régional de Bourgogne (N° 2012-9201AAO05S01353).

Résumé

Ce travail de thèse a pour but d'examiner l'émergence et l'évolution du caractère obligatoire des automatismes cognitifs. Pour satisfaire cet objectif, nous avons conçu une nouvelle situation expérimentale que nous avons appelée *Stroop musical*. Il s'agit d'une portée en clé de sol comprenant une note, présentée dans différentes positions, dans laquelle un nom de note, congruent ou incongruent avec la position, est écrit. Nous avons montré, à l'aide de ce paradigme, que les musiciens traitent plus lentement les noms de notes dans la condition incongruente que dans la condition congruente (Etude 1). Cet effet, nommé *effet Stroop musical* (MSE), est généré par l'automaticité de la dénomination de notes. Le Stroop musical offre la possibilité d'étudier l'évolution du caractère obligatoire de la dénomination de notes en évitant les biais liés à l'âge des sujets. Ainsi, nous avons testé plusieurs groupes d'enfants musiciens d'âge similaire dont le niveau de solfège variait de 1 à 5 ans. Nos résultats indiquent une relation linéaire positive entre le MSE et le niveau de pratique musicale (Etude 3), ce qui tend à montrer que le caractère obligatoire du traitement automatique augmente de façon monotone avec la pratique. En soumettant des musiciens adultes (Etude 2) et enfants (Etude 4) aux deux tâches conflictuelles du paradigme de Stroop musical, la lecture de mots et la dénomination de notes, nous avons également révélé que le pattern d'interférence dépend de la force relative des deux traitements en compétition. Enfin, nous avons constaté que l'automaticité de la dénomination de notes persiste malgré un arrêt total et prolongé de la pratique (Etude 5).

Mots clés : automatismes cognitifs, effet Stroop, interférence, pratique musicale, apprentissage, attention, enfants.

Abstract

The aim of this thesis is to examine the emergence and the evolution of the obligatory characteristic of cognitive automatisms. To achieve this, we devised a new experimental situation called *musical Stroop*. The basic arrangement comprises a treble staff with a note in various positions. A name of a note is printed inside the note. For the congruent condition, the note name is congruent with the note position on the staff, whereas in the incongruent condition, note name and position are incongruent. We showed that musicians process the incongruent condition slower than the congruent condition (Study 1). This effect, named *Musical Stroop Effect* (MSE), is generated by the automaticity of note naming. The musical Stroop offers the possibility to investigate the evolution of the irrepressibility of note naming, while avoiding bias related to subject age. Thus, we tested several groups of musician children of similar age whose the level of musical education varied from 1 to 5 years. Our results indicate a positive linear relation between the MSE and the level of musical training (Study 3). Consequently, the irrepressibility of the automatic processing seems to increase monotonically with practice (that is to say in parallel with the other characteristics of automatisms). We also showed, by submitting adults (Study 2) and children (Study 4) musicians to the two conflicting tasks of the musical Stroop paradigm, word reading and note naming, that the pattern of interference depends on the relative strength of the two competing processing. Finally, we noted that the automaticity of note naming persists despite a total and protracted cessation of practice (Study 5).

Keywords: cognitive automatisms, Stroop effect, interference, musical practice, learning, attention, children.

Liste des tableaux

Liste des figures

Liste des annexes

Table des matières

A la mémoire de Geneviève Bannier

« Car il ne faut pas se méconnaître : nous sommes automate autant qu'esprit ; et de là vient que l'instrument par lequel la persuasion se fait n'est pas la seule démonstration. Combien y a-t-il peu de choses démontrées! Les preuves ne convainquent que l'esprit. La coutume fait nos preuves les plus fortes et les plus crues ; elle incline l'automate qui entraîne l'esprit sans qu'il y pense. »

Pascal, Pensées (p. 1119-20)

INTRODUCTION GENERALE

Préambule

Le concept d'automaticité a fomenté la curiosité des chercheurs depuis les prémices de la psychologie expérimentale. Cattell (1886) fut le premier à opérer une distinction entre les processus automatiques et contrôlés. Il établit cette dichotomie après avoir constaté que la dénomination d'un objet ou de la couleur d'un patch était plus lente que la lecture d'un mot ou d'une lettre. Il expliqua ses résultats en avançant une interprétation qui porte une résonnance étonnamment moderne : « dans le cas des mots et des lettres, l'association entre l'idée et le nom est survenue tellement souvent que le processus est devenu automatique, alors que dans le cas des couleurs et des images nous devons choisir le nom par un effort volontaire » (p. 65). Les conclusions de Cattell influencèrent les travaux des psychologues de son époque (e.g., James, 1890, p. 559; Quantz, 1897) et réapparurent après l'avènement du cognitivisme (e.g., Shiffrin & Schneider, 1977).

James (1890), Jastrow (1906), et Wundt (1896/1897, 1903) font également partie des premiers auteurs à avoir présenté des descriptions de l'automaticité dont certaines idées furent reprises dans des versions contemporaines.

La notion d'automatisme ne prend sens que par contraste à l'égard des concepts de conscience, d'attention, et de contrôle volontaire. Le béhaviorisme, en écartant la conscience de son champ d'investigation (voir Watson, 1913), semblait donc orienter les psychologues vers l'étude de l'automaticité. Pourtant, les travaux relatifs aux automatismes furent paradoxalement délaissés par ce courant de pensée. Ce n'est qu'à partir des années 1950, au début de l'ère cognitiviste, que l'intérêt porté au traitement automatique de l'information a refait surface. Cette résurgence découle d'une évolution dans la manière d'appréhender le fonctionnement psychique, mais également du foisonnement des recherches menées sur l'attention. Certaines de ces études ont mis en évidence les limites intrinsèques des capacités de traitement du système cognitif et souligné la nécessité d'un mécanisme de sélection de l'information (e.g., Cherry, 1953). En conséquence, plusieurs théories postulèrent que l'attention serait une propriété structurelle qui agirait comme un filtre sélectif (e.g., Broadbent, 1958; Deutsch & Deutsch, 1963). Néanmoins, Atkinson et Shiffrin (1968) ont présenté un modèle de la mémoire dans lequel l'attention n'est plus envisagée comme un filtre mais comme un système de contrôle des informations. Dans le même ordre d'idée, Kahneman (1973) proposa un modèle fonctionnel de l'attention car il estimait que les théories de l'attention sélective ne prenaient pas en compte la quantité d'effort investie dans la focalisation attentionnelle. Il décrivit l'existence d'un processeur central dont le rôle serait d'affecter de l'attention sur les différents processus sollicités dans une situation donnée. Ce modèle suggère également que la pratique répétée et consistante d'un comportement entraîne une diminution des demandes attentionnelles. Les processus sont alors considérés comme automatiques dès qu'ils opèrent indépendamment des ressources attentionnelles. Le concept d'automaticité connaîtra par la suite d'importants développements grâce à des publications particulièrement influentes (e.g., Hasher & Zacks, 1979; Logan, 1988; Posner & Snyder, 1975; Shiffrin & Schneider, 1977).

Les travaux relatifs au traitement automatique apparaissent aujourd'hui dans des champs de recherche très diversifiés, aussi bien en linguistique et en biologie que dans tous les domaines de la psychologie. Par ailleurs, si les automatismes sont souvent mis en évidence par les inadaptations qu'ils entraînent, dans la vie quotidienne (e.g., Norman, 1981), comme en situation expérimentale (e.g., Stroop, 1935), il serait déraisonnable de les considérer avec dédain. Ils représentent en effet un avantage adaptatif inestimable lorsqu'ils sont exécutés dans une situation adéquate, car ils permettent de réaliser un traitement en sollicitant très peu de temps et de ressources, ce qui offre la possibilité de traiter d'autres informations en parallèle. Laberge et Samuels (1974) ont soutenu par exemple qu'il serait impossible d'accéder à la signification d'un texte écrit si l'identification des lettres et des mots ne pouvait se réaliser de manière automatique. Les automatismes sont nécessaires à toute activité complexe, et certains auteurs ont même argué qu'ils étaient la base habituelle de l'expérience consciente (e.g., Tzelgov, 1997).

L'étendue des recherches sur les automatismes et l'importance qu'ils revêtent dans nos comportements journaliers semblent suggérer que le concept d'automaticité fait l'objet d'une définition claire et consensuelle. A plus ample examen, toutefois, il apparaît que de nombreuses questions restent à élucider. Plusieurs études remettent notamment en cause l'idée selon laquelle l'évolution des différentes propriétés des automatismes en fonction de la pratique serait parallèle (e.g., Schadler & Thissen, 1981; Tzelgov, Henik, & Leiser, 1990; Tzelgov & Kadosh, 2009). Il s'avère en effet que le caractère obligatoire du traitement automatique (qui renvoie au fait qu'un comportement puisse se déclencher hors de toute intention, et être difficile à interrompre une fois initié) n'évoluerait pas de façon croissante et monotone avec la pratique. Ainsi, une pratique prolongée, en permettant d'atteindre un certain niveau d'expertise, mènerait à une récupération du contrôle sur le traitement automatique. Cependant, les recherches qui ont investigué cette question se sont heurtées à des problèmes méthodologiques (que nous analyserons en détail par la suite) difficilement surmontables, et elles ne peuvent, par conséquent, apporter de réponses satisfaisantes.

L'objectif principal de notre travail est d'étudier l'émergence et l'évolution du caractère obligatoire du traitement automatique en s'affranchissant des biais rencontrés dans les expériences qui se sont focalisées sur ce sujet. Pour satisfaire cette ambition, nous avons commencé par concevoir un paradigme original que nous avons appelé *Stroop musical*. Il s'agit d'une portée musicale comprenant une note, présentée à différentes positions, dans laquelle un nom de note (congruent ou incongruent) est inscrit. Ce cadre méthodologique a permis de mettre en évidence le caractère obligatoire de la dénomination de notes chez les musiciens (Etude 1). Afin d'exploiter les possibilités offertes par cette nouvelle situation expérimentale, nous nous sommes intéressés à la façon dont interagissent deux traitements hautement automatisés (la lecture de mot et la dénomination de notes) dans les deux tâches potentielles du Stroop musical (la tâche de lecture de mots et la tâche de dénomination de notes ; Etude 2). Puis, nous avons examiné l'évolution du caractère obligatoire de la dénomination de notes chez des enfants musiciens en cours d'apprentissage du solfège (Etude 3). Nous avons approfondi cette étude en soumettant de nouveaux enfants musiciens aux deux tâches conflictuelles du Stroop musical (Etude 4). Enfin, nous avons réalisé une dernière étude pour évaluer le devenir de processus, qui ont été automatisés dans le passé, après un arrêt total et prolongé de la pratique de l'activité (Etude 5).

Nous présenterons dans un premier temps le cadre théorique nécessaire à la compréhension des questions qui ont motivé les travaux menés dans cette thèse. Puis, nous exposerons successivement les cinq études que nous avons préalablement énoncées. Les résultats essentiels et leurs implications théoriques et pratiques seront discutés.

Introduction

Les automatismes cognitifs sont habituellement définis et diagnostiqués à l'aide des propriétés qui les caractérisent. Nous commencerons ainsi par recenser les traits assignés au traitement automatique et à décrire leur principal mode d'opérationnalisation. Nous aborderons ensuite les conceptions qui rendent compte de l'acquisition des automatismes. Dans une troisième partie, nous évoquerons les questions laissées en suspens par les théories de l'automatisation et les divergences présentes dans la littérature à propos de l'évolution des propriétés du traitement automatique. Pour terminer, nous examinerons les études qui se sont penchées sur ces discordances à l'aide du paradigme de Stroop, et nous proposerons une méthode qui serait susceptible d'apporter des réponses plus convaincantes que celles mises en évidence jusqu'alors.

1. Les caractéristiques des automatismes cognitifs

Deux propriétés principales sont communément attribuées au traitement automatique : l'absence de charge mentale et l'absence de contrôle intentionnel (Perruchet, 1988). D'autres caractéristiques, dont l'importance semble moins substantielle, sont évoquées de manière épisodique. Dans un souci d'efficacité, nous présenterons essentiellement les propriétés qui semblent recueillir les faveurs de la majeure partie des auteurs. Nous éluderons par conséquent les débats qui portent sur des caractéristiques secondaires et sur lesquels les avis sont plus partagés (pour un exposé plus exhaustif, voir Moors & De Houwer, 2006).

Il convient également de préciser que nous nous intéressons aux automatismes acquis par la répétition intensive d'une activité finalisée ; ce qui exclut les automatismes innés et les comportements acquis de manière incidente. Logan (1992) proposa d'ailleurs de réserver le terme *automatique* exclusivement aux processus automatiques appris intentionnellement.

1.1. L'absence de charge mentale

L'absence de charge mentale se caractérise par le fait qu'un comportement automatique ne nécessite aucun coût cognitif pour être exécuté, c'est-à-dire qu'il peut être accompli sans perturber la réalisation d'un autre traitement (automatique ou contrôlé) : on dit alors que les traitements opèrent en parallèle. Le terme « efficace » est parfois employé dans la littérature (e.g., Moors & De Houwer, 2006). Il peut sembler plus approprié dans la mesure où il désigne un processus qui ne consomme pas *ou peu* de ressources attentionnelles. Il rend ainsi mieux compte du caractère graduel et relatif du traitement automatique (cf. partie 2.1, intra).

Expérimentalement, on peut mettre en évidence cette propriété dans des situations de double tâche. L'objectif est d'adjoindre à une tâche primaire dont on veut évaluer le degré d'automaticité, une tâche secondaire qui doit être effectuée simultanément avec la première sans affecter sa réalisation. Par exemple, Spelke, Hirst, et Neisser (1976) ont demandé à des sujets de lire des histoires courtes (qu'ils devaient comprendre) tout en écrivant des mots sous dictée. Au début de l'expérience, le temps de lecture des participants était considérablement allongé en double tâche. Mais après 6 semaines d'entraînement, ils lisaient (en comprenant ce qu'ils lisaient) tout aussi rapidement en double tâche qu'en simple tâche. Le niveau d'expertise en lecture s'est ainsi accentué au cours de l'expérience, et les sujets sont devenus capables de réaliser cette tâche en parallèle avec la seconde et sans sacrifier leur compréhension.

1.2. L'absence de contrôle intentionnel

L'absence de contrôle intentionnel, également désigné par les termes obligatoire ou irrépressible, concerne le fait qu'un comportement automatique puisse se déclencher de façon involontaire, et être difficile à interrompre une fois initié. Ce second critère peut être mis en évidence dans les situations où un comportement automatique est non-pertinent par rapport au contexte dans lequel il apparaît, comme dans le célèbre paradigme de Stroop. Il s'agit d'une tâche qui consiste à dénommer la couleur avec laquelle un mot de couleur est écrit (e.g., le mot « bleu » écrit en rouge). Lorsque les deux dimensions sont en compétition (le mot et la couleur), le temps de réponse (TR) observé est alors plus important que dans une situation non-conflictuelle (i.d., lorsque les stimuli n'ont pas de signification ; il peut s'agir, par exemple, de dénommer la couleur d'un non-mot, comme « fdvpa » écrit en rouge, ou d'une série de Xs colorés). En effet, dans une situation conflictuelle, les sujets ne peuvent pas s'empêcher de lire le mot (alors que la tâche ne requiert pas qu'ils le fassent ; la consigne spécifie clairement que l'information verbale doit être ignorée). Cela entraîne ainsi une augmentation du TR par rapport à une situation non-conflictuelle. L'interférence désigne la différence entre le TR de la situation conflictuelle (i.e., la condition incongruente) et le TR de la situation non-conflictuelle (i.e., la condition neutre), et reflète le caractère obligatoire du traitement automatique mis en jeu par la lecture d'un mot.

1.3. Les autres propriétés

Parmi les autres traits associés à la notion d'automatisme figure la vitesse d'exécution, et le fait que le comportement automatique est résistant à des facteurs potentiellement perturbateurs tels que le vieillissement, le stress, la dépression (Hasher & Zacks, 1979), ou encore l'alcool (Birnbaum, Taylor, Johnson, & Raye, 1987).

Une autre propriété, plus contestée, est celle d'inconscience. Elle se manifeste par une incapacité à verbaliser la nature d'un processus ou d'un événement, car un comportement automatique ne mène pas à une mise en mémoire efficace de l'information sur laquelle il opère. Néanmoins, bien que de nombreux chercheurs aient associé l'automaticité à la notion d'inconscience (e.g., Jacoby, 1991; Kihlstrom, 1987; Posner & Snyder, 1975), certains auteurs considèrent qu'il ne s'agit pas d'une caractéristique nécessaire du traitement automatique (e.g., Bargh, 1992; Tzelgov, 1997).

2. Les conceptions de l'automaticité

2.1. La perspective dualiste

La première conception qui a prévalu dans la période cognitiviste estimait que les traits mentionnés précédemment se manifestaient en ''tout ou rien'' (e.g., Shiffrin & Schneider, 1977). Selon cette perspective, la pratique d'une activité entraînerait une réduction progressive des exigences attentionnelles, et le comportement serait considéré comme automatique lorsqu'il deviendrait indépendant de l'attention (Hasher & Zacks, 1979). Cette conception, que nous appellerons dualiste, admet qu'un processus est soit automatique, soit non-automatique. Chacun des deux modes de traitement serait composé d'un ensemble fixe de propriétés. Les caractéristiques de chacun de ces modes seraient donc parfaitement corrélées. Ainsi, on pourrait déterminer la nature d'un processus en investiguant uniquement la présence ou l'absence d'une seule caractéristique. Cette perspective a cependant été récusée par des études révélant un manque de co-occurrence entre les propriétés du traitement automatique (e.g., Kahneman & Treisman, 1984; Logan, 1989). De plus, certains travaux ont montré que l'automaticité n'était pas nécessairement indépendante de l'attention.

Une étude réalisée par Kahneman et Henik (1981) permet d'illustrer ce dernier point. Ces auteurs ont conçu une situation expérimentale de type Stroop qui consistait à présenter simultanément deux mots, un mot neutre et un mot de couleur, l'un au-dessus de l'autre. Les sujets devaient dénommer la couleur avec laquelle un des deux mots était présenté (l'autre mot était présenté en noir). La position du mot pertinent variait de manière imprédictible entre les essais. D'après la conception dualiste, l'automaticité de la lecture devrait entraîner le traitement des deux mots présentés quelle que soit la condition expérimentale (que le mot coloré soit le mot neutre ou le mot de couleur). Par conséquent, l'interférence observée devrait être similaire dans les deux conditions. Néanmoins, les auteurs ont constaté une interférence d'environ 200 ms lorsque le mot coloré était le mot de couleur, alors qu'elle s'avérait relativement faible lorsque le mot coloré était le mot neutre. Ces résultats ne pouvaient pas s'expliquer par la position rétinienne ou le mouvement des yeux car la présentation des items a été réalisée avec un tachistoscope (pendant un temps très bref) et les deux mots apparaissaient à la même distance de la fovéa. Ainsi, dans cette situation, l'acte de lecture dépend de la focalisation attentionnelle. L'information verbale, dont le traitement est tenu pour automatique, peut donc être négligée si l'attention n'est pas orientée vers elle. Francolini et Egeth (1980) ont également mis en évidence des résultats analogues.

Par ailleurs, d'autres études indiquent que la conjonction entre les propriétés attribuées au traitement automatique n'est pas toujours respectée. Par exemple, Kahneman et Chajczyk (1983) ont créé une variante du paradigme de Stroop dans laquelle le nom d'une couleur est écrit en noir au-dessus ou en-dessous d'une plage colorée. Cette plage apparaît pendant 200 ms à l'endroit du point de fixation, et la tâche du sujet est de dénommer sa couleur le plus rapidement et le plus précisément possible. L'interférence observée dans cette situation se trouve réduite de moitié si un autre mot, neutre, est simultanément présenté au sujet, du côté opposé au nom de couleur. Ainsi, dans cette situation, la lecture est à la fois obligatoire et consommatrice de ressources, ce qui contredit radicalement la conception dualiste. Paap et Ogden (1981), ainsi que Regan (1981), sont parvenus à la même conclusion en montrant qu'un traitement peut requérir un coût cognitif sans nécessiter de contrôle intentionnel. Les deux propriétés principales du traitement automatique ne sont donc pas nécessairement combinées.

La dissociation entre les principaux indicateurs de l'automaticité remet largement en cause la perspective dualiste. De plus, il apparaît clairement que les propriétés du traitement automatique ne représentent pas un caractère absolu. L'absence de contrôle intentionnel semble tout à fait relative (Francolini & Egeth, 1980; Kahneman & Henik, 1981). De même, la lecture des mots, habituellement considérée comme automatique, ne peut s'exercer totalement en parallèle (Kahneman & Chajczyk, 1983). Par ailleurs, la conception dualiste est dépourvue d'un mécanisme d'apprentissage explicite qui expliquerait le développement de l'automaticité (Logan, 1991, 1992). Cette perspective ne spécifie pas pourquoi un processus aurait besoin de moins de ressources attentionnelles après la pratique.

En vertu de toutes ces critiques, certains auteurs ont conclu que le concept d'automaticité manquait de consistance interne et devrait être abandonné (e.g., Regan, 1981). Néanmoins, Logan (1985) apporta une alternative à cette position. Il supposa que l'automaticité d'un processus évoluerait de manière graduelle et serait déterminée par la quantité de pratique. Il suggéra également que chaque propriété pourrait évoluer à un rythme particulier avec l'entraînement (p. 373). Le caractère obligatoire d'un processus pourrait survenir avec peu de pratique, alors que l'efficacité (i.e., le fait qu'un processus sollicite peu de ressources cognitives) apparaîtrait seulement après un exercice intensif.

L'idée d'un continuum de l'automaticité est particulièrement pertinente dans l'étude de MacLeod et Dunbar (1988). Ces auteurs ont assigné un nom de couleur spécifique à quatre formes géométriques non-familières (polygones). Ils ont ensuite demandé aux sujets d'apprendre ces noms au cours de plusieurs sessions de pratique. Après un entraînement de quantité variable, les formes étaient présentées en couleur de manière congruente ou incongruente avec leur nom. Les participants devaient réaliser deux tâches : dénommer la forme en ignorant la couleur, et dénommer la couleur sans tenir compte de la forme. Au début de l'entraînement (une journée), la couleur avait un effet sur la dénomination de la forme mais la forme ne produisait pas d'effet sur la dénomination de la couleur. Une pratique plus soutenue (5 jours) conduisait chaque dimension à interférer avec l'autre de façon équivalente. Après un entraînement intensif (20 jours), la forme avait un effet sur la dénomination de la couleur mais la couleur n'avait plus d'effet sur la dénomination de la forme. Le pattern de résultats était donc totalement opposé à celui obtenu après une journée de pratique. Ainsi, le degré d'automaticité de la dénomination de la forme semblait augmenter progressivement au cours de l'entraînement. En début de pratique, la dénomination de la couleur était plus automatisée que la dénomination de la forme. Après 5 jours, le niveau d'automaticité des deux traitements était à peu près équivalent. Enfin, après 20 jours, la dénomination de la forme était nettement plus automatisée que la dénomination de couleur. L'idée d'un continuum de l'automaticité semble donc parfaitement cohérente avec les résultats obtenus.

Suite aux objections dont la perspective dualiste a été l'objet, deux grandes conceptions se sont développées pour rendre compte de l'acquisition des automatismes.

2.2. Le renforcement d'une procédure algorithmique

Une première perspective conçoit que le renforcement d'une procédure algorithmique constituerait le principal mécanisme de l'automatisation (e.g., Anderson, 1992; Rosenbloom & Newell, 1986). Tout automatisme commencerait par l'exécution délibérée d'une procédure avec le concours de l'attention, mais la répétition intensive de l'activité entraînerait un retrait progressif de l'attention et du contrôle intentionnel du déroulement des opérations. L'automatisation résulterait ainsi du changement d'un même processus sous-jacent. L'exécution d'un même algorithme, qui représenterait dans un premier temps un traitement non-automatique, deviendrait avec la pratique un comportement automatisé. L'algorithme serait donc accompli plus rapidement et plus efficacement avec l'entraînement. Cette conception ne fait cependant pas l'unanimité.

2.3. La récupération directe en mémoire

Logan (1988) proposa une perspective selon laquelle l'automatisation procéderait d'un changement qualitatif dans la nature des opérations effectuées. Il expliqua le développement de l'automaticité comme la transition d'un traitement de nature algorithmique à une récupération directe en mémoire. Cette théorie suppose qu'un novice commencerait à utiliser un algorithme pour réaliser une nouvelle tâche. En répétant la même activité, il apprendrait des solutions spécifiques pour répondre à des problèmes spécifiques. Ainsi, il pourrait récupérer ces solutions lorsqu'il serait confronté de nouveau au même problème, ou réutiliser un algorithme approprié. Après une pratique intensive, la récupération en mémoire deviendrait plus efficace et plus rapide que le traitement algorithmique. L'apprenant abandonnerait alors complètement cette dernière procédure. C'est à partir de ce moment que le traitement serait considéré comme automatique.

Prenons l'exemple du calcul mental pour illustrer cette conception. Si on demande à un enfant qui apprend les opérations arithmétiques élémentaires d'effectuer un calcul comme 3×4, il commencera certainement par appliquer un algorithme. Il calculera tout d'abord $4 + 4$ (en comptant éventuellement sur ses doigts), puis il additionnera de nouveau 4 au résultat de la première opération pour arriver au résultat final : 12. Mais si cet enfant est confronté fréquemment à la même opération, il retiendra $3 \times 4 = 12$. Il sera alors en mesure donner le résultat de cette opération (3×4) sans effectuer aucun calcul, mais en récupérant directement la solution en mémoire.

Selon Logan (1988), l'acquisition d'un automatisme n'est pas due à la diminution des ressources attentionnelles allouées à la réalisation d'une tâche mais résulte d'un accès direct et rapide à une représentation mnésique unique adaptée à la situation. Dans cette perspective, un traitement automatique solliciterait peu d'attention car l'accès en mémoire est plus rapide et moins coûteux que l'élaboration d'une procédure algorithmique. De la même façon, la perte de contrôle serait consécutive au fait que la récupération directe en mémoire est largement déterminée par le contexte, sans évitement possible.

Cette théorie permet également de rendre compte du caractère graduel de l'automatisation. Si nous considérons en effet qu'une activité complexe met en jeu de nombreuses composantes de traitement, le passage d'une procédure algorithmique à la récupération directe en mémoire pourrait s'opérer à différents moments pour chacune de ces composantes. De ce point de vue, la perspective dualiste de l'automaticité n'a plus de pertinence.

Dans les deux conceptions, la récupération directe en mémoire et le renforcement d'une procédure algorithmique, la consistance de la pratique est un facteur essentiel pour le développement de l'automaticité. Néanmoins, les deux perspectives impliquent différents types de consistance (Carlson & Lundy, 1992). Un changement vers une récupération directe en mémoire se manifeste après une pratique sur des données consistantes, alors que le renforcement d'une procédure algorithmique requiert que les algorithmes restent consistants avec la pratique, mais pas les données sur lesquelles ils opèrent. En se basant sur les architectures mnésiques qui distinguent les systèmes de mémoire déclarative et procédurale, certains auteurs ont suggéré que les algorithmes qui avaient atteint un degré d'automaticité suffisant pourraient être stockés en mémoire procédurale (e.g., Anderson, 1992, 1996; Tzelgov, Yehene, & Naveh-Benjamin, 1997).

La conception relative au renforcement d'une procédure algorithmique est soutenue par des études sur le développement des habiletés dans lesquelles les bénéfices de la pratique d'algorithmes consistants sont transférés à des données qui n'ont jamais été rencontrées (e.g., Carlson & Lundy, 1992; Schneider & Fisk, 1984; Smith & Lerner, 1986). Mais certains auteurs ont avancé que les deux mécanismes, le renforcement algorithmique et la récupération directe en mémoire, pourraient sous-tendre le développement de l'automaticité (Tzelgov, Henik, Sneg, & Baruch, 1996; Tzelgov, Yehene, Kotler, & Alon; 2000).

3. Les questions non-résolues

3.1. Problème de diagnostique

Les deux grandes conceptions relatives au développement graduel de l'automaticité laissent néanmoins de nombreuses questions en suspens. Tout d'abord, elles ne donnent pas d'indications sur la manière de distinguer empiriquement les processus automatiques et non-automatiques. La perspective dualiste estime qu'un processus devient automatique dès lors qu'il opère indépendamment des ressources attentionnelles. Il est alors considéré comme un état terminal, insensible à la pratique (Hasher & Zacks, 1979). D'après l'approche graduelle, un comportement automatique n'est pas figé, et son évolution suit un continuum (e.g., MacLeod & Dunbar, 1988). Mais à partir de quel niveau de pratique peut-on attribuer le statut d'automaticité à un traitement ? Quel critère empirique faudrait-il satisfaire dans la mesure où il n'existe pas de ligne de démarcation précise entre les deux types de processus (automatiques et non-automatiques) ? Bargh (1992) proposa une solution à ce problème en considérant la présence d'une caractéristique particulière comme un critère minimal de l'automaticité : l'autonomie. Selon lui, un processus est autonome lorsqu'une fois initié, il peut s'accomplir entièrement sans contrôle conscient. Cependant, le choix de ce critère semble relativement arbitraire, et certains auteurs recommandent en conséquence de diagnostiquer les propriétés de l'automaticité individuellement en se gardant de toute généralisation (e.g., Moors & De Houwer, 2006).

3.2. Les divergences d'évolution des propriétés

Mais un problème plus substantiel semble compromettre la validité du concept d'automaticité. Il apparaît en effet qu'il existerait des divergences d'évolution des différentes propriétés des automatismes au cours de la pratique, ce qui serait contraire à l'idée d'une convergence entre les critères communément admise dans la littérature.

Il semble bien établi que la pratique d'une tâche tende à réduire progressivement la charge mentale nécessaire à son exécution (Perruchet, 1988). Lien, Allen, Ruthruff, Grabbe, McCann, et Remington (2006) ont indiqué, par exemple, que le degré d'automaticité de la lecture était plus marqué chez des personnes âgées (71 ans de moyenne d'âge) que chez des jeunes adultes étudiants à l'université (25 ans de moyenne d'âge). Dans cette étude, une tâche de décision lexicale était combinée avec une tâche visuelle ou auditive. Les performances observées révélaient que les personnes âgées étaient capables de réaliser les deux tâches en parallèle, mais pas les jeunes adultes. Les auteurs en ont conclu que l'automaticité de la lecture s'accentuait au cours de la vie, ce qui résulterait d'une plus grande expérience avec le traitement lexical. Il est également avéré que la vitesse d'exécution d'une tâche augmente de façon monotone avec la pratique. Crossman (1959) a montré, dans le domaine des habiletés, que la vitesse de fabrication artisanale de cigares continuait à augmenter après sept ans de pratique.

Néanmoins, s'il paraît manifeste que l'efficacité et la vitesse d'exécution évoluent de façon monotone avec la pratique, le problème semble plus complexe à propos de l'absence de contrôle intentionnel. De nombreuses observations tendent à montrer que le manque de contrôle atteindrait rapidement un maximum, pour ensuite décroître progressivement (e.g., Schadler & Thissen, 1981). Pourtant, l'automaticité et le contrôle ont souvent été considérés comme deux concepts opposés (e.g., Schneider, Dumais, & Shiffrin, 1984). Certains travaux concernant les ratés de la vie quotidienne indiquent cependant que la persistance d'une habitude ancienne ne conduit pas nécessairement à accentuer le caractère obligatoire du comportement appris (e.g., Reason, 1984). Au contraire, l'inadaptation de l'action peut provenir de la perturbation d'anciennes routines par la généralisation d'un comportement acquis récemment. Ces résultats pourraient donc remettre en cause l'idée selon laquelle l'évolution des différentes propriétés des automatismes serait parallèle.

3.3. Un concept à revoir ?

Ces données poseraient un problème de définition évident et rendraient le diagnostic de l'automaticité plus difficile à établir. De plus, les divergences d'évolution des différents critères des automatismes ne sont pas compatibles avec la notion d'opérateur central auquel on attribue les propriétés constitutives de l'attention et d'un centre de contrôle. Par conséquent, s'il est avéré que l'évolution du caractère obligatoire des automatismes n'est pas croissante et monotone, cela conduirait à revoir la définition du concept d'automaticité ainsi que les propriétés que l'on attribue à l'opérateur central.

Ces résultats, bien qu'intriguant, s'accordent pourtant avec les observations concernant la notion d'habileté. Logan (1985) signala en effet que des sujets possédant une habileté très développée étaient davantage capables de la contrôler que des sujets disposant de la même habileté à un niveau d'automaticité plus réduit. Il indique par exemple que des dactylographes expérimentés peuvent inhiber leur frappe sur un clavier suite à un signal d'erreur de façon très rapide. Il en conclut que « la croyance générale que l'automaticité et le contrôle sont opposés peut être erronée, ou au moins, exagérée » (p. 379).

Nous allons maintenant nous attacher à recenser les études qui ont tenté de caractériser la nature de la relation entre le contrôle et le niveau de pratique d'un traitement. Le cadre méthodologique privilégié pour investiguer le degré de contrôle d'un traitement est le paradigme de Stroop. De nombreuses études se sont intéressées à la manière dont évoluait l'effet Stroop (qui reflète l'obligation de lire un mot écrit) en fonction du niveau de pratique de la lecture. Nous ferons une revue de ces travaux afin de déterminer ce qu'ils nous apportent sur l'évolution du caractère obligatoire du traitement automatique.

4. L'évolution du caractère obligatoire en fonction du niveau de pratique

4.1. L'évolution de l'effet Stroop en fonction de l'âge

4.1.1. Avec la procédure par essais multiples

Comalli, Wapner, et Werner (1962) furent les premiers à étudier l'évolution de l'interférence en fonction de l'âge dans une tâche de type Stroop. Ils ont présenté aux sujets une carte composée de mots de couleurs (RED, BLUE, et GREEN[1]) imprimés avec une encre de couleur incongruente à la signification du mot (condition incongruente). La tâche des participants consistait à dénommer le plus correctement et le plus rapidement possible la couleur avec laquelle les items étaient imprimés. Les performances observées dans cette situation furent comparées à une condition neutre. Il s'agissait d'une carte comprenant des patchs rectangulaires imprimés en couleur. Sur chaque carte (i.e., dans les deux conditions), trois couleurs ont été utilisées : rouge, bleu et vert. Les résultats indiquaient que l'interférence (i.e., la différence du TR moyen entre la condition incongruente et la condition neutre) était très marquée à l'âge de 7 ans, et qu'elle diminuait progressivement au cours de l'enfance, jusqu'à 17-19 ans. Puis, elle restait relativement stable à l'âge adulte, et amplifiait sensiblement à partir de 65 ans.

Schiller (1966) réalisa une étude similaire à celle de Comalli et al. (1962), mais il s'intéressa uniquement à des enfants scolarisés de la première année de primaire jusqu'à l'université. Cependant, les sujets du premier groupe (i.e., des enfants en première année de primaire) étaient moins âgés que les plus jeunes participants de Comalli et al. Les résultats

[1] Toutes les études que nous présenterons avec le paradigme de Stroop (utilisant des stimuli verbaux) ont été réalisées en anglais, excepté les travaux de Faccioli, Peru, Rubini, et Tassinari (2008), et Peru, Faccioli, et Tassinari (2006), réalisés en italien. Lorsqu'un autre langage fut employé, l'information est précisée.

montraient que l'interférence s'avérait très faible en première année de primaire, qu'elle devenait maximale en deuxième année puis déclinait graduellement jusqu'à l'université. Ainsi, après une brève phase de croissance, l'interférence s'atténuait au fur et à mesure que le niveau scolaire augmentait.

Dash et Dash (1982) répliquent les résultats de Schiller (1966) avec cinq groupes d'enfants scolarisés de la première à la cinquième année de primaire (un groupe de sujets par année). Ils observent une relation en U inversé entre l'interférence et le niveau scolaire avec un maximum atteint en deuxième année. Dans une étude longitudinale réalisée sur trois ans, de la deuxième à la quatrième année de primaire, les auteurs constatèrent à nouveau que l'interférence était maximale en deuxième année et qu'elle diminuait jusqu'en quatrième année (Expérience 2).

Plus récemment, Armengol (2002) réutilisa le même matériel que Comalli et al. (1962) en espagnol, et elle compara des enfants provenant d'une école privée à des enfants issus d'une école publique. Tous les participants étaient scolarisés en primaire (de la première à la sixième année). Un groupe de sujets fut constitué pour chaque âge (de 6 à 12 ans) et chaque école. L'interférence observée pour les enfants de l'école publique était relativement faible à 6 ans, augmentait jusqu'à l'âge de 7 ans, et diminuait progressivement par la suite. Pour les enfants de l'école privée, l'interférence était maximale à 6 ans et s'atténuait graduellement avec l'âge. Le pattern de résultats obtenu pour les erreurs était similaire à celui des TRs. L'auteur explique les différences de performances entre les deux écoles par de meilleures capacités de lecture pour les enfants scolarisés en école privé. Leur compétence dans cette activité était suffisamment avancée dès la première année d'enseignement pour que l'interférence soit maximale. Les résultats d'un test de vocabulaire (une adaptation en espagnol de l'échelle de vocabulaire en images Peabody) réalisé avec chaque participant permettaient de confirmer cette hypothèse. L'école privée possédait davantage de ressources, dispensait plus de cours (deux heures supplémentaires par jour), et la formation des professeurs était plus aboutie que dans l'école publique.[2] De plus, les enfants de l'école privée provenaient de familles dont les moyens financiers étaient beaucoup plus importants que celles des enfants issus de l'école publique (le niveau d'instruction des parents d'élèves provenant de l'école privée était aussi plus élevé). Leur environnement scolaire et culturel paraissait donc plus favorable à leur épanouissement intellectuel, et c'est certainement la

[2] L'étude a été réalisée dans la ville de Mexico.

raison pour laquelle leurs habiletés de lecture étaient plus marquées dès la première année de primaire.

Il semblerait ainsi que l'interférence obtenue dans une tâche de Stroop augmente de façon sensible en début d'apprentissage de la lecture, atteigne rapidement un maximum (après un à deux ans de pratique), et diminue progressivement lorsque le comportement devient plus automatisé. Néanmoins, dans toutes les études précédentes, les items ne sont pas présentés individuellement. Tous les stimuli d'une condition donnée sont inscrits sur une carte que le sujet doit traiter le plus rapidement possible (nous parlerons de procédure par essais multiples). Il existe deux catégories d'erreurs, dont la quantité varie en fonction de l'âge, dans ce type de procédure (voir Rand, Wapner, Werner, & McFarland, 1963) : les réponses déviantes sur les items (e.g., des réponses mal articulées, comme le fait d'énoncer ''blau'' au lieu de ''blue''), et les réponses déviantes sur la séquence (e.g., l'insertion de mots entre les réponses, comme ''yes'' ou ''and''). La présentation individuelle des stimuli (ou procédure par essais distincts) permet d'éviter cette dernière catégorie d'erreurs. Par ailleurs, la procédure par essais multiples ne sollicite pas les mêmes processus perceptifs. Dulaney et Rogers (1994) ont mentionné par exemple que cette méthode requiert un balayage systématique qui n'est pas nécessaire lorsque les items sont présentés individuellement. De plus, la sensibilité à la distraction visuelle et les processus perceptifs inhérents à la procédure par essais multiples pourraient varier en fonction de l'âge (Davidson, Zacks, & Williams, 2003). La présentation individuelle des items est exempte de ces aspects perceptifs, et elle présente l'avantage d'être une méthode informatisée, ce qui permet d'obtenir des TRs plus précis. Il est également probable qu'elle soit plus sensible pour estimer l'effet d'interférence (voir Faccioli, Peru, Rubini, & Tassinari, 2008, p. 281) ; le fait de mixer les items congruents et incongruents pouvant, par exemple, minimiser la possibilité d'utiliser des stratégies pour répondre.

4.1.2. Avec la procédure par essais distincts

Peru, Faccioli, et Tassinari (2006) ont mis au point une procédure par essais distincts dans une tâche de Stroop couleur-mot afin d'examiner l'évolution de l'interférence chez des participants âgés de 6 à 10 ans. Leur population comprenait également deux groupes d'enfants pré-lecteurs de 3 et 5 ans. Les résultats indiquaient qu'un effet d'interférence significatif était présent chez les sujets de 3 ans et qu'il augmentait légèrement à 5 ans.[3] Ensuite, l'interférence amplifiait très fortement chez les enfants de 6 ans, atteignait un maximum à 7 ans, et diminuait graduellement jusqu'à 10 ans. Les auteurs ont rapporté que la tendance quadratique entre l'âge et l'interférence était significative.

Ainsi, la relation en U inversé observée entre l'interférence et le niveau de lecture semble indépendante de la procédure utilisée car elle est aussi obtenue avec une présentation individuelle des stimuli. Ce résultat paraît toutefois relativement énigmatique. En effet, l'interférence obtenue dans une tâche de Stroop est censée refléter le degré d'automaticité de la lecture (Logan, 1997), et plusieurs études indiquent que l'automaticité de la lecture s'accentue de manière graduelle avec la pratique (e.g., Lien et al., 2006; Ruthruff, Allen, Lien, & Grabbe, 2008). Par conséquent, l'interférence devrait être corrélée positivement avec les habiletés de lecture. Mais les données empiriques vont à l'encontre de cette hypothèse. Il semblerait même qu'après une augmentation en début d'apprentissage, l'interférence diminue en fonction de la pratique.

4.1.3. Avec le paradigme de Stroop image-mot

Cette relation négative entre l'interférence et le niveau de lecture a également été obtenue avec une tâche de Stroop image-mot (e.g., Guttentag & Haith, 1979; Rosinski, Golinkoff, & Kukish, 1975). Cette situation expérimentale, qui est une variante de la tâche de Stroop classique, consiste à présenter des dessins d'objets familiers sur chacun desquels un mot est écrit. Les participants sont tenus de dénommer les images le plus rapidement et le plus précisément possible sans tenir compte des informations verbales. Rosinski et al. (1975) ont constaté, en utilisant cette méthode, que le temps de dénomination des objets était plus important dans la condition incongruente (i.e., quand les mots et les images sont incongruents mais correspondent à la même catégorie sémantique, comme par exemple, le mot ''Pig'' inscrit sur le dessin d'un chat) que dans la condition congruente (i.e., quand les mots correspondent aux images, comme le mot ''Gun'' écrit sur le dessin d'un pistolet) pour leur trois groupes de sujets. Cet effet de la congruence était relativement prononcé chez les enfants de deuxième année de primaire, diminuait en sixième année, et devenait encore plus réduit chez les adultes. Rosinski (1977) observa un pattern de résultats analogues avec un

[3] Selon les auteurs, l'interférence obtenue avec les enfants de 3 ans proviendrait d'un effet perceptif, alors que chez les enfants de 5 ans, elle serait due, au moins en partie, à certaines connaissances verbales (car les enfants sont exposés fréquemment à du matériel verbal dans leur environnement). Les deux groupes de pré-lecteurs ne semblaient pas capables de traiter le sens des mots. En effet, aucune différence n'était observée entre la condition incongruente et la condition congruente dans chacun de ces groupes.

groupe de participants supplémentaire (des enfants en quatrième année d'école primaire). Cependant, dans les deux études précédentes, lorsque la condition incongruente était comparée à une condition neutre (dans laquelle des non-mots prononçables étaient écrits sur les images), l'interférence ne variait pas en fonction du niveau scolaire. Il est toutefois possible que la condition neutre adoptée ne soit pas la plus appropriée. En effet, les non-mots utilisés rimaient avec les mots de la catégorie sémantique des images sur lesquelles ils étaient présentés (e.g., le non-mot ''Pog'', qui rime avec ''Dog'', était inscrit sur le dessin d'un chat). Ainsi, les non-mots choisis ont pu entraîner un effet d'interférence dans une condition supposée neutre. De manière générale, le problème concernant le choix d'une ligne de base adéquate dans une tâche de type Stroop n'a jamais été résolu (voir, e.g., Neely, 1991; Melara & Algom, 2003; Sabri, Melara, & Algom, 2001).

Une autre situation de type Stroop, n'impliquant pas la lecture, permet d'évaluer l'interférence en fonction du niveau de pratique de la dimension non-pertinente. Il s'agit des tâches qui font intervenir le traitement automatique de la valeur numérique. Cela peut consister, par exemple, à présenter deux chiffres qui diffèrent à la fois par leur valeur et par leur taille (de police). Le sujet doit alors déterminer lequel des deux chiffres est le plus grand numériquement ou physiquement en fonction de la tâche demandée. Girelli, Lucangeli, et Butterworth (2000) ont montré une relation en U inversé (dans une tâche de comparaison physique de chiffres) entre l'âge et l'interférence avec un maximum atteint en cinquième année de primaire.

4.1.4. Conclusions

Il existe néanmoins un problème commun à toutes ces études : le degré d'automaticité du traitement interférent est largement confondu avec l'âge des participants. En effet, les auteurs ont constitué leurs groupes de sujets en fonction de l'âge (e.g., Armengol, 2002; Comalli et al., 1962; Girelli et al., 2000; Peru et al., 2006) ou du niveau scolaire (qui est également lié à l'âge; e.g., Dash et Dash, 1982; Guttentag & Haith, 1979; Rosinski, 1977; Rosinski et al., 1975; Schiller, 1966). De cette façon, chaque groupe était caractérisé par un certain degré d'automaticité de la lecture ou du traitement de la valeur numérique. Ainsi, l'évolution de l'interférence peut aussi bien être expliquée par les habiletés à traiter la dimension non-pertinente que par des facteurs liés à l'âge des participants, comme le niveau de maturation neurale et cognitive. Certains auteurs ont même avancé que la relation en U inversé observée entre l'interférence et l'âge dans une tâche de type Stroop résultait de deux tendances opposées mais concomitantes (Cohen, Dunbar, & McClelland, 1990) : une augmentation du degré d'automaticité du traitement interférent, et une augmentation de la maturation des fonctions exécutives et attentionnelles basées sur les circuits neuronaux situés dans le cortex préfrontal et le cortex cingulaire antérieur (Adleman et al., 2002; Carter, Mintun, & Cohen, 1995; Zysset, Müller, Lohmann, & von Cramon, 2001). L'interférence augmenterait tant que le premier processus prévaut sur le second, après quoi, elle diminuerait (malgré des habiletés de lecture, ou numériques, plus développées) : le traitement de la dimension non-pertinente deviendrait de plus en plus irrépressible, mais dans le même temps, l'habileté à inhiber la réponse correspondant aux mots écrits ou aux chiffres s'améliorerait. L'accroissement de l'interférence observée de nouveau à partir de 60 ans (Comalli et al., 1962; voir aussi Roelofs & Hagoort, 2002) pourrait s'expliquer de la même façon par une détérioration des fonctions exécutives et attentionnelles. Afin de déterminer précisément la manière dont évolue le caractère obligatoire du traitement automatique en fonction du degré de pratique, il semblerait donc opportun de se focaliser sur le niveau de lecture ou des habiletés numériques indépendamment de l'âge des sujets.

4.2. L'évolution de l'effet Stroop en fonction du niveau de lecture

4.2.1. Chez des sujets sans difficultés de lecture

Plusieurs études ont tenté d'étudier l'évolution de l'interférence en fonction des habiletés de lecture. Schadler et Thissen (1981) ont par exemple constitué six groupes de sujets en fonction de leurs performances à un test de lecture. Les résultats montraient que l'interférence était absente pour le groupe de non-lecteurs. Puis, elle commençait à apparaître chez les lecteurs débutants, amplifiait lorsque les habiletés de lecture augmentaient (jusqu'à un niveau équivalent à celui habituellement observé en quatrième année de primaire), et déclinait chez les lecteurs les plus avancés. Mais la confusion avec l'âge restait présente dans cette étude ; l'âge moyen des groupes s'avérait d'autant plus élevé que les capacités de lecture étaient importantes. La manipulation du degré de pratique de la lecture semble donc limitée par des contraintes éthiques et pratiques qu'il est difficile d'éviter.

Malgré tout, la diminution de l'interférence est également observée quand le contrôle du niveau de lecture est précisément établi. Il n'est pas certain, toutefois, que cette évolution reflète réellement une propriété fondamentale des automatismes. En effet, Schadler et Thissen (1981) ont expliqué la réduction de l'interférence par le fait que les enfants les plus âgés posséderaient un meilleur contrôle sur les processus cognitifs que les enfants débutants. Cette plus grande capacité de contrôle permettrait aux enfants plus âgés, soit de plus facilement ignorer la dimension interférente (i.e., le mot), soit de traiter la dimension pertinente (i.e., la couleur) et la dimension interférente du stimulus de manière indépendante (p. 140). Evaluer l'influence de la pratique en éliminant l'effet potentiel de l'âge semble donc impossible avec le Stroop traditionnel, étant donnée la très forte liaison entre l'âge et le niveau de lecture.

D'autres études ont essayé de contourner le problème en se focalisant sur des participants avec des difficultés de lecture.

4.2.2. Chez des sujets avec difficultés de lecture

Ehri (1976) réalisa une tâche de Stroop image-mot avec quatre groupes de sujets : des enfants de deuxième et cinquième année de primaire, des lycéens, et des étudiants (respectivement âgés de 8.3, 11.3, 14.7, et 22.7 ans en moyenne). Elle compara une situation incongruente à une situation neutre dans laquelle une image seule était présentée. Les résultats montraient une diminution graduelle de l'interférence en fonction du niveau scolaire. Néanmoins, l'auteur ajouta un groupe d'enfants mauvais lecteurs de deuxième année de primaire (8.2 ans de moyenne) et constata qu'ils n'étaient pas perturbés par la présence des mots incongruents pour dénommer les images. Leur temps de dénomination dans cette situation ne différait pas de celui obtenu dans la condition neutre. Il semble donc que l'interférence augmente de manière importante en début d'apprentissage de la lecture indépendamment de l'âge (la décroissance de l'interférence reste difficile à interpréter car elle est confondue avec l'âge), ce qui a été confirmé par Stanovich, Cunningham, et West (1981). Ces derniers ont réalisé une étude longitudinale dans laquelle des enfants scolarisés en première année de primaire effectuaient une tâche de Stroop couleur-mot à trois reprises (en septembre, février, et avril). L'interférence amplifiait fortement entre la fin du mois de septembre et le milieu du mois de février, mais l'augmentation devenait nettement plus faible entre mi-février et fin avril.

Ehri et Wilce (1979) ont également rapporté que l'interférence s'accentuait en début d'apprentissage de la lecture, et diminuait par la suite lorsque le degré d'automaticité de cette activité augmentait, en étudiant deux groupes de sujets d'un âge sensiblement équivalent. Ces auteurs ont entraîné les participants à lire les mots interférents utilisés dans une tâche de Stroop image-mot. A partir d'un échantillon d'enfants scolarisés en première et deuxième année de primaire, ils composèrent deux groupes de sujets sur la base d'un pré-test qui consistait à lire une liste de 20 mots le plus rapidement possible (en passant les mots que les participants ignoraient). Les enfants qui identifiaient au moins 16 mots furent placés dans un groupe que nous appellerons bons lecteurs. Tous les autres formèrent un second groupe que nous qualifierons de mauvais lecteurs.[4] Les sujets étaient ensuite entraînés à lire la liste de mots (ils devaient identifier puis déterminer une fonction pour chacun des mots afin de s'assurer qu'ils connaissaient leur signification). Tous ces items étaient utilisés comme noms interférents dans une tâche de Stroop image-mot réalisée avant et après l'entraînement. Les résultats ont mis en évidence une augmentation de l'interférence lors de la seconde passation chez les mauvais lecteurs et une diminution chez les bons lecteurs. Ehri et Wilce montrèrent que la réduction de l'interférence chez les bons lecteurs n'était pas une conséquence de la répétition de la tâche (voir Dyer, 1971). En effet, lorsque la même étude fut répliquée sans entraînement de lecture, l'interférence ne variait pas entre les deux passations de la tâche de Stroop pour les deux groupes de sujets (constitués de la même façon que dans l'expérience précédente, avec d'autres enfants; Expérience 3).

Les divergences d'évolution de l'interférence entre les bons et les mauvais lecteurs seraient liées, selon les auteurs, à des effets d'apprentissage différents entre les deux groupes. L'entraînement conduirait les mauvais lecteurs à traiter davantage de mots de manière obligatoire, ce qui produirait en conséquence une plus grande interférence. Concernant les bons lecteurs, l'entraînement les amènerait à traiter les mots plus rapidement et plus efficacement (les mots étaient déjà traités sans contrôle intentionnel avant l'entraînement), ce qui aurait pour effet de réduire l'interférence.

Des résultats semblables ont été obtenus par Pace et Golinkoff (1976), bien qu'ils ne se soient pas focalisés spécifiquement sur l'évolution de l'interférence en fonction du niveau d'automaticité de la lecture.

4.2.3. Chez des sujets dyslexiques

De nombreuses études ont comparé l'interférence observée chez des sujets dyslexiques dans des tâches de type Stroop par rapport à des participants du même âge sans difficultés de lecture. Par exemple, Everatt, Warner, Miles, et Thompson (1997) ont montré que des enfants dyslexiques (âgés de 10 à 11 ans en moyenne) avaient un effet d'interférence similaire à des sujets plus jeunes d'un niveau de lecture comparable, mais plus forte que des enfants du même âge sans problèmes de lecture. Les dyslexiques seraient moins capables, selon les

[4] L'âge moyen des groupes n'est pas précisé, mais les auteurs ont indiqué que les 16 bons lecteurs étaient en première année de primaire, et parmi les 20 mauvais lecteurs, 6 étaient en deuxième année de primaire et le reste en première.

auteurs, de contrôler le traitement du mot (et notamment de l'inhiber). De la même manière, Helland et Asbjørnsen (2000) ont constaté que l'interférence était plus grande chez des enfants dyslexiques (environ 12 ans) que chez des enfants sans difficultés de lecture du même âge (voir aussi Kapoula et al., 2010, pour des résultats analogues avec des adolescents). Ils expliquèrent cette différence par une altération des fonctions exécutives chez les dyslexiques.

Protopapas, Archonti, et Skaloumbakas (2007) ont présenté aux sujets de leurs expériences deux feuilles comportant, pour la première, un ensemble d'items constitués de Xs colorés, et pour la seconde, des mots de couleurs imprimés avec une couleur incongruente. Ils comparèrent des enfants d'une école publique (en septième année de scolarisation) avec un échantillon clinique du même âge (12.5 ans en moyenne) atteint de dyslexie. L'interférence observée était plus importante pour les enfants dyslexiques que pour les enfants sans difficultés de lecture. Les auteurs rapportèrent également, en s'intéressant à un groupe d'enfants tout venant (non-dyslexiques; 12.6 ans de moyenne d'âge), que l'interférence était corrélée négativement avec les habiletés de lecture, mais pas avec l'âge des participants (Expérience 2).

Faccioli et al. (2008) ont évalué l'interférence dans une tâche de Stroop couleur-mot avec des enfants dyslexiques en utilisant une procédure par essais distincts. Quatre groupes de sujets scolarisés de la deuxième à la cinquième année d'école primaire sont comparés avec des enfants sans problèmes de lecture et du même niveau scolaire. Les résultats indiquent que l'interférence est toujours plus forte pour les enfants dyslexiques. Cependant, elle ne diffère pas significativement entre les quatre groupes. Hicks et Jackson (2005) ont également investigué l'évolution de l'interférence avec des sujets dyslexiques (âgés de 8 à 30 ans), mais ils ont constitué cinq groupes de sujets selon leur niveau de lecture (sans préciser toutefois l'âge moyen de chacun des groupes). Il s'avérait que l'interférence était corrélée négativement avec le niveau de lecture. Cependant, ces données sont délicates à interpréter ; le manque d'informations sur l'âge moyen des groupes nous convie à la circonspection.

4.2.4. Conclusions

De manière générale, les dyslexiques, malgré des habiletés de lecture moins développées (et donc moins automatisées), ont une interférence plus grande que des sujets du même âge sans difficultés de lecture. Ainsi, le contrôle cognitif de la dimension à ignorer semble plus efficient avec des lecteurs plus expérimentés, ce qui est incompatible avec l'idée d'une évolution croissante et monotone du caractère obligatoire du traitement automatique.

Mais il est malaisé de déterminer si les performances des dyslexiques (comme celles des mauvais lecteurs ; voir, par exemple, Ehri & Wilce, 1979) sont dues principalement à des caractéristiques intrinsèques de l'automaticité (notamment un plus faible degré d'automatisation) ou à des troubles des fonctions exécutives et attentionnelles comme certains auteurs le suggèrent (e.g., Helland & Asbjørnsen, 2000).

Les résultats observés en comparant des sujets (d'un âge donné) selon leurs habiletés de lecture semblent toutefois cohérents avec ceux obtenus en se focalisant sur l'âge des participants. Il apparaît en effet que l'interférence se manifesterait très tôt et augmenterait fortement en début d'apprentissage de la lecture (e.g., Ehri, 1976; Stanovich et al., 1981). Elle diminuerait par la suite (après un à deux ans de pratique) quand le degré d'automaticité de la lecture deviendrait plus élevé (e.g., Ehri & Wilce, 1979). Néanmoins, l'interprétation de ces résultats reste difficile à établir. La décroissance de l'interférence pouvant être expliquée par des facteurs (notamment exécutifs et attentionnels) indépendants de l'automaticité. De plus, les données empiriques ne convergent pas toutes vers une relation en U inversé entre l'interférence et le niveau de pratique du traitement non pertinent. Martin (1978) a mis en évidence une relation positive entre les habiletés de lecture et l'interférence dans une tâche de Stroop couleur-mot. Parmi un panel d'adultes âgés de 18 à 30 ans, l'auteur a composé deux groupes de sujets à partir d'un test de lecture. Il s'agissait de lire une liste de mots aléatoires aussi vite et aussi précisément que possible. La moitié des participants les plus rapides constituait le premier groupe (lecteurs rapides), l'autre moitié le second (lecteurs lents). Les résultats révélaient que les lecteurs rapides avaient une plus grande interférence que les lecteurs lents (expérience 2). Par ailleurs, Rousselle et Noël (2007) n'ont constaté aucune différence significative, dans une tâche de comparaison physique de chiffres de type Stroop, entre des sujets avec et sans difficultés du traitement numérique (tous les participants étaient en deuxième année de primaire). L'effet de la congruence était similaire pour les deux groupes. Ces divergences nous amènent à considérer une autre approche pour répondre à notre question théorique.

4.3. Le paradigme de Stroop chez les bilingues

Une méthode plus prometteuse, à première vue, pour étudier l'évolution du caractère obligatoire du traitement automatique en fonction du niveau de pratique, consiste à utiliser le paradigme de Stroop standard avec des mots appartenant à une seconde langue. L'acquisition d'une seconde langue survient plus tard que l'apprentissage de la lecture, parfois même à l'âge adulte (et donc à une période du développement où le système nerveux et les fonctions cognitives sont matures), et implique une pratique intensive et naturelle qui peut être relativement bien contrôlée. Cette méthode peut ainsi permettre d'évaluer l'interférence en fonction du degré d'automaticité de la lecture du second langage en évitant les biais liés à l'âge (chez des sujets sans problèmes cognitifs).

4.3.1. Comparaison des effets Stroop obtenus dans les deux langages

Cette situation expérimentale a essentiellement été utilisée pour investiguer la structure de la représentation des connaissances chez les bilingues, notamment les relations lexicales et sémantiques entre les différents langages (e.g., Altarriba & Mathis, 1997). Les études concernant le contrôle du traitement automatique (avec ce paradigme spécifique) ne sont pas légion, mais certains travaux ont apporté des informations notables sur l'automatisation. Tzelgov et al. (1990) ont montré par exemple que le contrôle cognitif était plus efficace pour la première langue (L1) que pour la seconde (L2). Ils ont manipulé la fréquence du langage des mots écrits (à ignorer). Les sujets étaient informés de cette manipulation. Ils savaient ainsi que la plupart des mots présentés seraient en L1 ou en L2 (80% des mots étaient dans un langage, 20% dans l'autre). Les auteurs considéraient que les attentes des participants (à propos d'un langage donné) induiraient un plus grand contrôle sur le traitement (des mots écrits attendus). L'effet de la congruence (seules une condition congruente et une condition incongruente furent utilisées) s'atténuait de façon significative quand les mots attendus appartenaient principalement à L1, quel que soit le langage de réponse. Aucune différence n'était observée (l'effet de la congruence ne variait pas) quand les mots attendus concernaient essentiellement L2. Ce résultat fut obtenu avec des participants de L1 arabe et de L2 hébreu (Expérience 1), et de la même manière, pour des sujets de L1 hébreu et de L2 arabe (Expérience 2). Les auteurs ont avancé que des mécanismes de contrôle appliqués aux processus de lecture seraient mis en œuvre à partir d'un certain degré d'automaticité, ce qui expliquerait la relation en U inversé observée dans les études qui évaluent l'interférence en fonction du niveau de pratique de la lecture (p. 769).

Néanmoins, certains résultats semblent contredire l'idée que l'interférence diminue (après une brève phase de croissance) quand l'automaticité d'un traitement s'accentue. En effet, si cette relation était vérifiée, l'interférence devrait être plus importante avec des mots présentés en L2 qu'avec des mots présentés en L1. Sumiya et Healy (2008) ont pourtant constaté (avec des participants adultes de langue maternelle anglaise qui apprenaient le japonais) que l'interférence apparaissait plus forte quand les mots étaient présentés en L1 (quel que soit le langage de réponse). Ils ont également mentionné que la taille de l'interférence, pour les stimuli appartenant au second langage (réponses en L1), n'était pas significativement corrélée avec les habiletés de lecture en L2. Lee et Chan (2000) ont trouvé cependant que l'interférence était similaire dans les deux langues des sujets (des chinois adultes qui apprenaient l'anglais). Ils précisèrent toutefois que le niveau des participants dans les deux langues était équivalent. Ce dernier résultat suggère que l'interférence dépendrait du degré d'automaticité du traitement non-pertinent.

4.3.2. Evolution de l'automaticité du second langage

Peu d'études ont mesuré l'évolution de l'interférence en fonction du niveau de pratique avec des mots du second langage (en demandant aux participants de dénommer la couleur des stimuli dans leur langue maternelle). La première fut réalisée par Mägiste (1984) avec des participants de langue maternelle allemande (âgés de 14 à 19 ans) qui apprenaient le suédois. Elle utilisa deux tâches de type Stroop (couleur-mot et image-mot). L'interférence n'évoluait pas en fonction du niveau de suédois (évalué par la durée de résidence en Suède ; tous les sujets vivaient en Suède) dans la tâche de Stroop couleur-mot, alors qu'elle semblait augmenter de façon croissante dans la tâche de Stroop image-mot (aucune analyse inférentielle n'a été réalisée sur ces données). L'auteur suggéra que la quantité d'interférence induite par les mots suédois était directement liée à la pratique de ce langage. Chen et Ho (1986) ont observé une diminution de l'interférence à partir de la deuxième année de primaire et à nouveau une augmentation à l'université avec des participants chinois de seconde langue anglaise (l'acquisition de l'anglais commençait à l'école primaire). Mais l'âge était confondu avec le niveau de lecture.

Plus récemment, Braet, Noppe, Wagemans, et Op de Beeck (2011) ont réalisé une étude avec des participants adultes de langue maternelle allemande qui apprenaient le japonais (tous les sujets ont commencé à apprendre le japonais à l'université et ont suivi le même enseignement). Ils ont utilisé un mode de réponse manuel pour éviter les confusions entre les langages (les sujets devaient appuyer le plus rapidement possible sur une touche colorée qui correspondait à la couleur avec laquelle l'item était présenté). L'effet Stroop observé (TRs incongruent – TRs congruent) était plus fort en L1 qu'en L2. De plus, les auteurs ont montré une corrélation entre les deux effets Stroop (allemand et japonais) qui semblait suggérer que des différences individuelles affectaient les effets obtenus dans les deux langages. Ils ont donc calculé un rapport ΔStroop à partir de l'effet Stroop mesuré dans les deux langages tel que ΔStroop = Stroop(L2) / Stroop(L1) de manière à normaliser les scores individuels. Le principal résultat de cette étude concernait la corrélation positive entre le ΔStroop et le niveau de lecture des participants en japonais.

4.3.3. Conclusions

Les travaux sur les bilingues n'apportent pas de conclusions univoques sur la relation entre l'interférence Stroop et le degré d'automaticité du traitement non-pertinent. Bien que l'étude probante de Braet et al. (2011) indique une relation positive, peu de données empiriques ont répliqué cette liaison. L'utilisation du paradigme de Stroop chez les bilingues apparaît cependant problématique car de nombreuses questions à propos du bilinguisme sont encore non-résolues. La nature des processus automatisés avec la pratique d'une seconde langue est notamment débattue. Est-ce qu'il s'agit du lien entre les mots de chaque langage (e.g., RED-rouge, avec rouge activant RED et RED activant une réponse interférente[5]) ou de l'accès direct aux concepts (avec rouge activant directement une réponse interférente)? La réponse à cette question dépend d'options théoriques (voir en particulier le modèle d'association de mots vs. le modèle de médiation conceptuelle de Kroll et Stewart, 1994). A un niveau plus élémentaire, quel est le rôle de l'apprentissage de la correspondance graphème/phonème ? Ce rôle est certainement nul quand les deux langages partagent la plupart de leur alphabet (e.g., anglais/français), mais la réponse est moins claire lorsque les deux langages impliquent des symboles différents, comme l'anglais et le chinois (voir, par exemple, Chen & Ho, 1986). Il est probable que le poids de tous ces facteurs change avec la pratique, ce qui rend l'étude d'un éventuel processus d'automatisation relativement complexe. Quoi qu'il en soit, il semble exagéré de considérer que l'utilisation du paradigme de Stroop avec un second langage permet de reproduire ce qui survient initialement avec le premier langage, avec l'avantage de découpler l'âge et le niveau d'automaticité.

4.4. Un nouveau paradigme

Au regard des travaux qui ont porté sur l'évolution de l'interférence en fonction du degré de pratique du traitement à ignorer dans des tâches de type Stroop, il apparaît que de nombreuses difficultés entravent la mise en évidence de conclusions convaincantes. Le niveau de pratique de la lecture (ou plus généralement du traitement interférent) est intimement lié à

[5] En supposant une situation avec des bilingues anglais-français de langue maternelle anglaise.

l'âge des sujets, ce qui rend la manipulation de cette variable particulièrement délicate. Les études concernant les participants qui souffrent d'un défaut d'habiletés de la dimension interférente (comme les dyslexiques) semblent également discutables. Peut-on considérer, par exemple, qu'un enfant dyslexique scolarisé en cinquième année de primaire se distingue d'un sujet sans difficultés cognitives, du même âge et du même niveau scolaire, uniquement par un plus faible degré d'automaticité de la lecture ? Il est probable que des facteurs d'ordre exécutif et attentionnel rendent plus complexe la comparaison entre ces deux individus. Enfin, les études sur le bilinguisme sont aussi sujettes à caution, car il n'existe pas de consensus relatif à la nature des processus en cours d'automatisation lors de l'apprentissage d'une seconde langue.

Les principaux résultats semblent toutefois indiquer une relation en U inversé entre l'interférence et le niveau de pratique de la dimension non-pertinente (e.g., Schadler & Thissen, 1981; Schiller, 1966), mais plusieurs études tendent à montrer que cette relation pourrait être positive (e.g., Braet et al., 2011; Martin, 1978). Les données de MacLeod et Dunbar (1988; cf partie 2.1, intra) révélaient également une liaison positive entre l'interférence et le niveau de pratique du traitement à ignorer (avec des sujets adultes). En effet, l'interférence observée dans la tâche de dénomination de la couleur s'avérait d'autant plus grande que l'entraînement à dénommer les formes était important (Expérience 3). Notons par ailleurs que cette relation positive apparaissait indépendamment de l'âge, comme dans l'étude de Braet et al. (voir aussi Martin, 1978). D'autres travaux n'ont constaté en revanche aucune relation entre l'interférence et le degré d'automaticité de la dimension non-pertinente (e.g., Sumiya & Healy, 2008; Rousselle & Noël, 2007).

Le manque de consistance des résultats et les difficultés méthodologiques rencontrées dans les études qui se sont intéressées à l'évolution de l'interférence en fonction du niveau de pratique du traitement à ignorer dans les tâches de type Stroop nous ont conduits à envisager une nouvelle situation expérimentale pour répondre à la problématique de notre travail. Nous proposons en effet d'exploiter un paradigme alternatif que nous appellerons *Stroop musical*. Il s'agit d'une portée en clé de sol comprenant une note, présentée dans différentes positions, dans laquelle un nom de note est écrit. Dans la condition congruente, le nom de note est congruent avec la position de la note sur la portée (Figure 1a), alors que dans la condition incongruente, le nom de note et la position sont incongruents (Figure 1b). Un *effet Stroop musical* serait mis en évidence par une altération du traitement des noms de note dans la condition incongruente par rapport à la condition congruente. Ce résultat attesterait du caractère obligatoire du traitement automatique de la dénomination de notes chez les musiciens.

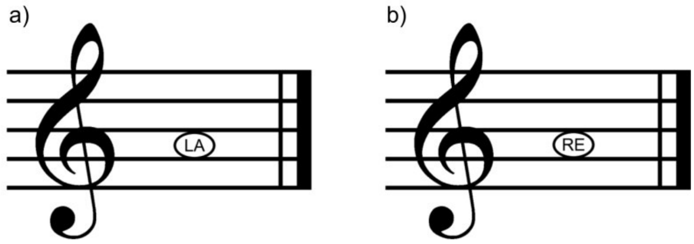

Figure 1. Exemples des différentes conditions : a) condition congruente; b) condition incongruente.

Ce paradigme comporte la plupart des avantages de la tâche de Stroop couleur-mot. Tout d'abord, la dénomination de notes est une activité naturelle, intensément pratiquée au cours de l'apprentissage du solfège, ce qui permet d'atteindre un niveau de pratique qu'il serait impossible d'obtenir dans une situation de laboratoire. Par ailleurs, la dénomination de notes, de même que la lecture, est la composante d'une activité plus complexe, dont l'automaticité est nécessaire à l'expression du comportement d'ensemble. Le traitement automatique des notes permet aux musiciens d'allouer de l'attention à des processus intégratifs consacrés à l'analyse des accords et des lignes mélodiques, et d'assurer également le contrôle moteur essentiel à la pratique d'un instrument.

L'utilisation du Stroop musical serait particulièrement appropriée pour évaluer le caractère obligatoire du traitement automatique de la lecture de notes chez les musiciens en fonction de leur niveau de pratique. En effet, contrairement aux autres tâches de type Stroop, cette situation autorise une plus grande flexibilité dans la manipulation de la pratique du traitement interférent, car l'apprentissage du solfège peut être plus facilement découplé de l'âge. Il est possible de trouver des enfants d'un âge donné qui ont un niveau de solfège différent, tout simplement parce qu'ils n'ont pas commencé l'enseignement de la musique au même âge. Ainsi, on pourrait constituer par exemple cinq groupes d'enfants âgés de 12 ans qui auraient de 1 à 5 ans de pratique musicale. Ce paradigme pourrait donc nous permettre d'éviter les biais liés à l'âge des participants et d'apporter ainsi une lumière nouvelle sur notre compréhension de l'automaticité.

La première étape de notre travail consistera à vérifier l'existence empirique de l'effet Stroop musical (MSE) avec des musiciens experts (Etude 1). Nous exploiterons ensuite les potentialités du Stroop musical en soumettant des musiciens adultes à deux tâches : la dénomination de notes et la lecture de mots. Ainsi, nous pourrons observer comment interagissent deux traitements automatisés après une pratique intensive et naturelle (la lecture et la dénomination de notes), ce qu'il n'est possible de réaliser avec aucune autre situation de type Stroop (Etude 2). Puis, nous nous focaliserons sur le point central de ce travail en étudiant l'évolution du caractère obligatoire de la dénomination de notes en fonction du niveau de pratique musicale. Pour cela, nous soumettrons les sujets au Stroop musical après avoir constitué différents groupes suivant leur niveau de solfège, et en contrôlant l'âge des participants (Etude 3). Nous approfondirons cette étude en présentant à de nouveaux sujets musiciens les deux tâches du Stroop musical pour observer l'évolution du MSE et du MSE inversé en fonction du degré d'automaticité de la dénomination de notes (Etude 4). Enfin, nous exposerons les résultats préliminaires de notre dernière étude qui concerne un aspect des automatismes qu'il est inconcevable d'investiguer avec la lecture. En effet, le Stroop musical offre une nouvelle possibilité d'exploration expérimentale. Il peut permettre d'évaluer ce qui reste d'un traitement qui a été hautement automatisé dans le passé après un arrêt prolongé de la pratique. L'avantage de la pratique musicale (et notamment du traitement de la hauteur des notes sur une portée), par rapport à la lecture, est qu'elle se montre plus facile à contrôler, car les notes de musiques ne sont pas omniprésentes dans notre environnement (comme le sont les mots). Nous pourrons ainsi observer si le caractère obligatoire de la dénomination des notes de musique persiste après une longue période sans pratique, et éventuellement, comment il évolue en fonction de la durée d'arrêt (Etude 5).

PARTIE EXPERIMENTALE

Etude 1

The Musical Stroop Effect:
Opening a new avenue to research on automatisms

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
CNRS-UMR 5022, University of Bourgogne, Dijon, France

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The Musical Stroop Effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Author Note

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat, Université de Bourgogne, LEAD/CNRS, UMR5022, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France

This research was supported by the grant ANR-09-JCJC-0129 and the grant Conseil Régional de Bourgogne N° 2012-9201AAO05S01353.

Correspondence concerning this article should be addressed to Laurent Grégoire, LEAD-CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 2100 Dijon, France. E-mail: laurent.gregoire@u-bourgogne.fr

Résumé

Cette étude examine si un effet de type Stroop pourrait être obtenu en remplaçant la lecture de mots par la dénomination de notes chez les musiciens. Pour ce faire, nous avons conçu une nouvelle situation expérimentale. Il s'agit d'une portée musicale avec une note, présentée dans différentes positions, dans laquelle un nom de note (congruent ou incongruent avec la position de la note) est écrit. Les résultats indiquent que le traitement des noms de note est significativement plus lent dans la condition incongruente que dans la condition congruente, aussi bien dans une tâche go/no-go (Expérience 1), que dans une tâche verbale (Expérience 2). Ce paradigme ouvre la voie à de nouvelles possibilités d'investigations sur les automatismes. Cette situation offre en effet une plus grande flexibilité dans la manipulation de la pratique (du traitement interférent) que la tâche de Stroop classique. Elle pourrait ainsi s'affranchir des difficultés rencontrées dans les études qui se sont penchées sur l'automatisation de la lecture (notamment les biais liés à l'âge des sujets). L'apprentissage du solfège n'est pas aussi régulé que l'apprentissage de la lecture, il serait donc plus aisé, par exemple, de trouver des enfants d'âge similaire mais ayant un niveau de pratique musicale différent (alors qu'il est très difficile de découpler l'âge du niveau de pratique de la lecture). Ce paradigme pourrait donc permettre d'étudier l'évolution du caractère obligatoire de la dénomination de notes chez les musiciens en fonction de leur niveau de pratique en évitant les biais liés à l'âge des participants. Par ailleurs, il devient également possible d'évaluer ce qui reste d'un traitement qui a été hautement automatisé dans le passé, après un arrêt prolongé de la pratique.

Abstract

The usual color-word Stroop task, as well as most other Stroop-like paradigms, has provided invaluable information on the automaticity of word reading. However, investigating automaticity through reading alone has inherent limitations. This study explored whether a Stroop-like effect could be obtained by replacing word reading with note naming in musicians. Note naming shares with word reading the crucial advantage of being intensively practiced over years by musicians, hence allowing to investigate levels of automatism that are out of reach of laboratory settings. But the situation provides much greater flexibility in manipulating practice. For instance, even though training in musical notation is often conducted in parallel with the acquisition of literacy skills during childhood, many exceptions make that it can be easily decoupled from age. Supporting the possibility of exploiting note naming as a new tool for investigating automatisms, musicians asked to process note names written inside note pictures in incongruent positions on a staff were significantly slowed down in both a go/no-go task (Experiment 1) and a verbal task (Experiment 2) with regard to a condition in which note names were printed inside note pictures in congruent positions.

Introduction

Stroop's (1935) classic article is certainly one of the most influential papers on cognition. The effect reported in this article is well known: Color naming is slowed down by an incongruent color word. The huge interest for this phenomenon stems from the fact that it allows exploring one of the major properties of automatisms, namely the fact that after extensive practice with a consistent task, the processes involved in this task tend to generate interference with other tasks in which they are in principle irrelevant.

A number of variants to the standard color-word version have been created, which were essentially devised to provide an alternative to the color dimension. In the picture-word paradigm, for instance, a word (e.g., "hand") is printed inside a congruent or an incongruent picture (e.g., a hand or a foot, respectively; e.g., Lupker, 1979), while in a spatial version of the Stroop task, the word "below" is displayed above a fixation point, or conversely (e.g., Palef & Olson, 1975). In both cases and many other Stroop-like paradigms (although not all; exceptions will be dealt with in the General Discussion), interference is generated by word reading.

Investigating word reading is of obvious interest, given the crucial importance of the issue in daily behavior. But focusing on reading is endowed with its own shortcomings. First, any generalization of conclusions drawn from a single experimental arrangement remains a matter of speculation. More importantly, reading offers quite limited opportunity to manipulate certain variables that would seem of primary interest when studying interference and automaticity. As claimed by MacLeod (1991), "practice may turn out to be one of the most effective manipulations for disentangling theories of the Stroop effect" (p. 182). Now, manipulating reading practice is severely restrained due to obvious practical and ethical constraints. For instance, the only possible approach to assess the amount of reading practice consists in using academic level or reading skill as an independent variable (e.g., Catling, Dent, Johnston, & Balding, 2010; Protopapas, Archonti, & Skaloumbakas, 2007). Unfortunately, reading ability is acquired within an age span where an overwhelming amount of other cognitive changes occurs, and the observed performance evolution may be due to multiple factors other than reading practice, if only the correct understanding of instructions by the youngest children.

In search for an alternative paradigm

In this paper, we propose to trade word reading for note naming in musicians, hence analyzing the possibility of observing what is coined herein as the *musical Stroop effect*. The basic arrangement comprises a staff with a note in various positions (see Figure 1). A name of a note is printed inside the note. For the congruent condition (Figure 1a), the note name is congruent with the note position on the staff, whereas in the incongruent condition (Figure 1b), note name and position are incongruent. A musical Stroop effect would be revealed by the impaired processing of the printed note name in incongruent conditions with regard to congruent conditions, an effect that would attest to the interference generated by note naming in musicians.

The musical Stroop would share most advantages of the classical color-word version to investigate automatisms. The most obvious is that note naming is an activity that is intensively practiced over years by musician experts, and relying on naturally occurring practice allows attaining a level of practice that is out of reach of laboratory settings. Moreover, note naming shares with word reading the nice property of being a component of more complex activities, the automaticity of which, far from generating anecdotic glitches, is required to ensure the successful expression of the whole behavior. The automaticity of note naming is necessary to allocate musicians' attention to higher integrative processes devoted to analyzing chords and melodic lines of the musical work and to ensure motor control for instrument playing, as the automaticity of identification of single words is needed to free our minds to deal with higher-order aspects of the task such as comprehension and metacognitive functions.

However, as an experimental means of investigation, the musical Stroop effect would avoid the shortcomings pointed out above, because musical formation provides much greater flexibility in manipulating practice than literacy learning. Control over the practice level is tighter, because music notes are not ubiquitously present, as words are, in our daily environment. Moreover, even though musical training is often conducted in parallel with the acquisition of basic literacy skills during childhood, there is a large number of exceptions, which makes that practice can be easily decoupled from age. For instance, instead of running experiments with children aged from 6 to 11 to investigate the first years of practice on word reading, exploiting the musical Stroop effect makes it possible to run experiments with 12-year-old children having from 1 to 6 years of musical practice. Conversely, it becomes possible to explore whether automatism formation interacts with age, by comparing the effect of extensive practice at various ages, even including elderly people (exploiting the fact that a proportion of newly retired persons start musical training). Among a host of other possibilities offered to experimental exploration, and without any possible analog in reading, is the fact that a number of persons having gained a high level of expertise in music turns out to give up any practice for years.

Is there previous evidence for a musical Stroop effect?

Zakay and Glicksohn (1985) exposed pianists to musical notes or to the names of those notes, which were written at congruent or incongruent locations on a staff. Participants performed verbal and manual responses in succession, in counterbalanced order. For the verbal response, they were required to name the notes according to their position on the staff while ignoring the printed name, or alternatively, to read aloud the printed names of the notes while ignoring their position on the staff. The manual response involved the same conditions, except that participants were required to press the appropriate piano key. Zakay and Glicksohn reported an overall effect of congruency. However, this effect interacted with the mode of response, and Zakay and Glicksohn did not assess the effect of congruency separately for each response mode (verbal vs. manual). More importantly, Zakay and Glicksohn did not distinguish the effect of congruency due to the automaticity of word reading (as virtually any other Stroop-like tasks) from the effect of congruency due to musical abilities, which is at focus in the present study.

Zakay and Glicksohn, nevertheless, provided detailed response times for each condition (see their Table 1, p. 418) and asserted, on the basis of a post-hoc Scheffé procedure, that a difference of 880 ms was required for any contrast to be significant at the .05 level.[6] Using this conservative benchmark, it appears that the musical Stroop effect did not reach significance: Incongruent conditions slowed down verbal responses, but to a far too limited extent (310 ms). The main effect of congruence was in fact driven by the conditions in which pianists had to respond through piano keys.[7] Although these early results raise some skepticism about the possibility of observing a musical Stroop effect as defined above, it

[6] This is the difference in RTs for a whole sequence of ten homogeneous stimuli.

[7] These data are potentially relevant for research on motor skills. However, a number of factors might have artificially inflated, if not entirely produced the congruence effect. For instance, when pianists have to press, say, a "SOL" written on the "LA" line in incongruent conditions, their response may be delayed by the interference due to the irrepressibility of the response "LA" (a Stroop-like effect), but the delay may also be due to the fact that the word "SOL" does not designate a single key on the piano keyboard. Subjects have to select arbitrarily at least between the two SOL keys surrounding the last pressed key, which one to press, and this selection is presumably time-consuming.

cannot be excluded that some features of the Zakay and Glicksohn's experiment, the description of which are postponed to the General Discussion, could have played down the effect measured with verbal responses.

The present study

The experiments below explored again the potential of using the automaticity of note naming in musicians instead of reading as a source of a Stroop-like effect. Our procedure is a conceptual replication of a subset of the conditions implemented in the Zakay and Glicksohn's (1985) study, with the addition of a number of controls and methodological improvements. The main experimental conditions (i.e., congruent and incongruent) were presented above (Figure 1a and 1b). In addition, words that were not names of notes were also printed inside the notes on some trials. In Experiment 1, both musicians and nonmusicians were told to press a key when the printed word was a note name, and to refrain any response when it was not. In Experiment 2, musicians and nonmusicians were instructed to read aloud printed words, while ignoring note positions. In both experiments, a musical Stroop effect should result in (1) longer response times in the incongruent than in the congruent condition for musicians and (2) an interaction of the congruity effect with musical expertise, attesting that the congruity effect is specific to musicians, and hence is likely to be a genuine consequence of musical practice.

A long-lasting debate in the Stroop literature is whether the Stroop effect, classically defined as the difference between performances in congruent and incongruent conditions, is mainly due to interference in the incongruent condition, facilitation in the congruent condition, or some mix of both effects. To explore this issue, a baseline situation was added, in which the note names were printed inside a picture of a note, but outside any contextual staff (Figure 1c).

Experiment 1

Both intuition and experimental data suggest that reading a word aloud may be less susceptible to interference than a more delayed response involving subsequent steps of processing. Earlier Stroop literature has shown that classifying the word as belonging to a predefined category (e.g., responding "animal" when the word is "frog" in a picture-word task, see Smith & Magee, 1980) increased the probability of interference with the irrelevant dimension. Independently, using a motor response instead of word reading (e.g., pressing color-labeled buttons in the color-word version, see Blais & Besner, 2006) is endowed with the same consequence. To increase our chance of getting an effect in the first experiment, these two methodological variants were combined. A go/no-go task was used (as Durgin, 2003), in which participants had to press a key only when the printed word was a note name. A selective slowdown in response time of musicians when the printed note name was incongruent with the note picture would attest to a musical Stroop effect.

Method

Participants. Twenty-eight volunteer undergraduate psychology students at the Université de Bourgogne received course credit for their participation. They were French native speakers and reported normal or corrected vision. Half of them had formal musical training and had played a musical instrument for at least five years (musicians: 8 women, 6 men). They had an average of 13.93 (*SD* = 3.77) years of musical training. Other participants had never studied or practiced music (nonmusicians: 9 women, 5 men).

Materials. For the main three experimental conditions, note names were printed inside note pictures, which could appear on each of the 13 possible positions going from C4 to A5. In the congruent- and incongruent-context conditions, the note pictures were presented in a treble staff (Figure 1a and 1b). For the congruent condition, the note name was congruent with the note position on the staff, whereas in the incongruent condition, note name and position were incongruent, with the name written inside the note picture being one of the six other possible note names.

In the out-of-context condition, the note pictures were presented without the staff, although for the sake of matching at best perceptual conditions, they appeared on the same 13 spatial locations, as if they were correctly positioned on a virtual staff (Figure 1c).

For the purpose of the go/no-go task, the to-be-read stimuli of two additional conditions were made up of words that were not names of notes. As the note names, they were six two-letter and one three-letter highly frequent French words (*CE, JE, TU, NI, TA, VU, and PAR*). These words were printed inside note pictures, and they were displayed at all 13 note locations either inside a staff (in-context condition) or without any surrounding staff (out-of-context condition).

Note names and non-note words appeared in standard uppercase font printed in black over a white background on a computer screen. The treble staff was 7.7 cm wide by 5.1 cm high. For each of the five conditions (congruent, incongruent, note names out-of-context, words in-context, and words out-of-context), the stimuli appeared six times on each of the 13 locations, leading to 78 trials per condition, and resulting in 390 trials (78 × 5) for the whole session.

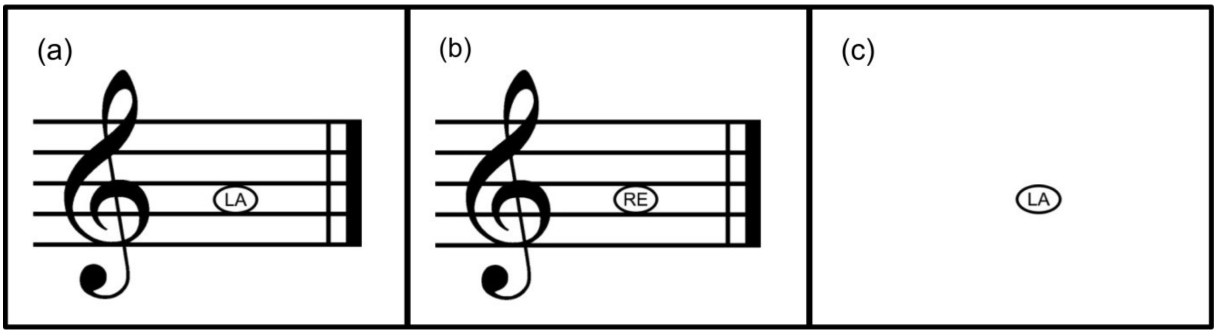

Figure 1. Examples of the congruent-context (a), the incongruent-context (b), and the out-of-context conditions (c). Note that in the musical French notation (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

Procedure. Participants were asked to press the space bar, as quickly as possible, when the printed word was a note name, and to refrain from responding when it was not. The next stimulus appeared after 1200 ms if no response had been made.
To prevent the iconic memory of the staff to influence the processing of the following note, the stimuli were randomly displayed at one of four possible positions without immediate repetition at the same location. The four positions were defined as the center of (invisible) rectangles resulting from the exhaustive partitioning of the screen into four quadrants of equal size. A fixation cross displayed for 1 s at the center of the screen preceded the stimulus, which stayed on the screen until participant's response. The inter-trial-interval was 1s. The 390 trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of note locations, note names, and non-note words. They were displayed as ten blocks of 39 trials each with a self-paced break between blocks. After the session, which lasted about 20 minutes, the musicians filled out a questionnaire about their musical abilities. This questionnaire included information about the number of years of musical training, which is reported above, and a lot of other questions that were not exploited, such as the instrument that was played (if any), the amount of daily practice, and whether participants possessed absolute pitch.

Results and Discussion

Values beyond three standard deviations of the mean (1.03%) were removed from the data for each participant. Misses and false alarms represented 1.04% and 3.37% of the trials, respectively.

Mean Response Times (RTs) for hits are shown in Figure 2. RTs were significantly longer in incongruent- than in congruent-context condition for musicians, with a mean difference of 9.91 ms (*SD* = 11.56), $t(13) = 3.21$, $p = .007$, $d = .86$. There was no corresponding effect of congruity for nonmusicians, $t(13) = 0.66$, $p = .52$. To assess whether the congruity effect significantly differed between the two groups, an ANOVA was carried out with Congruity (congruent-context, incongruent-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable. There was no main effect of congruity, $F(1, 26) = 2.47$, $p = .128$, a significant effect of musical expertise, $F(1, 26) = 33.11$, $p < .001$, $\eta_p^2 = .560$, and crucially a significant Congruity × Musical Expertise interaction, $F(1, 26) = 6.64$, $p = .016$, $\eta_p^2 = .203$.

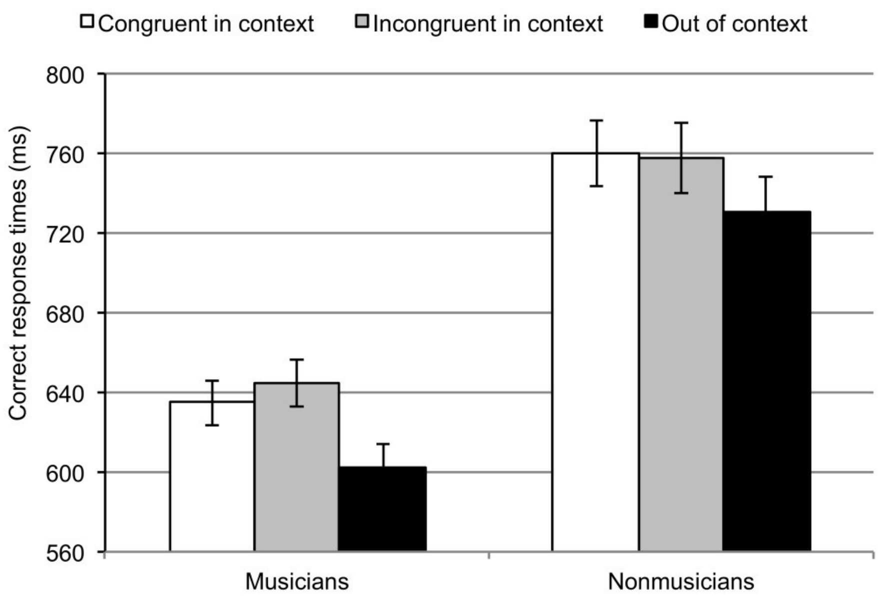

Figure 2. Correct response times as a function of Context and Musical Expertise in Experiment 1. Error bars indicate standard errors.

The out-of-context condition was devised to provide a baseline for assessing whether the Stroop effect observed in musicians was due to interference or facilitation. An ANOVA with Condition (congruent-context, incongruent-context, out-of-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable gave significant effects for condition, $F(2, 52) = 101.35$, $p < .001$, $\eta^2_p = .796$, musical expertise, $F(1, 26) = 35.142$, $p < .001$, $\eta^2_p = .575$, and Condition × Musical Expertise interaction, $F(2, 52) = 4.33$, $p = .018$, $\eta^2_p = .143$. RTs of musicians were shorter in the out-of-context condition than in the incongruent-context condition, $t(13) = 10.90$, $p < .001$, $d = 2.91$, but surprisingly, they were also shorter than in the congruent-context condition, $t(13) = 9.42$, $p < .001$, $d = 2.52$. This pattern suggests that even the congruent condition generates interference, a result that has sometimes been reported in the color-word version of the Stroop task (e.g., Nealis, 1973; Schulz, 1979). However, it turns out that the same effects occurred for nonmusicians. RTs of nonmusicians were also shorter in the out-of-context condition than in both the incongruent-context condition, $t(13) = 7.52$, $p < .001$, $d = 2.00$, and the congruent-context condition, $t(13) = 6.20$, $p < .001$, $d = 1.66$, suggesting that musical expertise does not play any role in this effect. Another interpretation for shorter RTs in the out-of-context condition stems from the degree of perceptual complexity of each condition. Word reading would be easier when the note is displayed alone than when it appears inside a complex frame staff.

If one assumes that musicians and nonmusicians are both sensitive to perceptual complexity, then the differences between musicians and nonmusicians in the congruent- and incongruent-context conditions can be compared to the difference between musicians and nonmusicians in the out-of-context condition. In the out-of-context condition, musicians were 127.35 ms faster than nonmusicians. The corresponding value for the congruent-context condition was nearly identical, $M = 125.12$, $F(1, 26) = 0.15$, $p = .706$, whereas the difference for the incongruent-context condition was significantly reduced, $M = 112.81$, $F(1, 26) = 7.59$, $p = .011$, $\eta^2_p = .226$. In other words, musicians and nonmusicians differed only in the incongruent condition once their difference in the out-of-context condition has been partialized out, which suggests that the musical Stroop effect is mainly due to interference in the incongruent condition.

To conclude, we obtained clear evidence for a musical Stroop effect using a go/no-go procedure: As anticipated, there was a reliable effect of congruity, and this effect was limited to musicians. On the other hand, data also revealed an unexpected main effect of musical expertise: In all conditions, musicians performed the go/no-go task considerably faster than nonmusicians. Recall that in the go/no-go task, participants had to decide whether the printed item was a note name or not. Although nonmusicians are familiar with note names, if only because they are part of many popular children's songs, it is very likely that the exhaustive set of note names is more accessible for musicians than for nonmusicians. A possibility is that a decision of categorical membership would require serial memory search in nonmusicians, while retrieval of the whole set of note names would be performed in parallel in musicians (Shiffrin & Schneider, 1977).

Experiment 2 aimed at addressing three remaining questions. First, as discussed above, the go/no-go task (a categorization task, which involves a motor mode of responding) cumulates two procedural changes known to make the required responses more susceptible to interference than reading aloud. Using this task does not invalidate or undermine our conclusion, but gives rise to our main question: Is a musical Stroop effect still present when the go/no-go task is replaced with more standard oral responses? Second, are the longer RTs observed in the in-context than in the out-of-context condition for both musicians and nonmusicians really due to the relative perceptual complexity of the respective displays? Third, is the unexpected overall slowness of nonmusicians due to the categorical membership decision required in the go/no-go task?

Experiment 2

Experiment 2 involved the very same five conditions as Experiment 1. However, participants had to read aloud the printed word (a note name or another word), while ignoring the position of the note. Reading times collected in the congruent and incongruent conditions should allow confirming evidence for a musical Stroop effect. Trials involving words that were not names of notes, which served only as foils in the go/no-go task of Experiment 1, were exploited in Experiment 2 to indicate whether the presence of the staff context is sufficient to slow down the responses. Because the notion of congruity is objectless for these items, any difference in RTs between in-context and out-of-context conditions would reveal the influence of a variable unrelated to musical information *per se*, presumably perceptual complexity. This information should allow assessing whether the RTs observed in the out-of-context condition involving notes as items can serve as a baseline for teasing apart inhibitory and facilitatory components of the musical Stroop effect. Finally, because a categorical membership decision is no longer involved, Experiment 2 should allow deciding whether the overall advantage of musicians over nonmusicians observed in Experiment 1 was task specific.

Method

Participants. Thirty-four new participants were recruited from the same population as in Experiment 1. Half of them had formal musical training and had played a musical instrument for at least five years (musicians: 11 women, 6 men). They had an average of 10.12 (*SD* = 3.86) years of musical training. Other participants had never studied or practiced music (nonmusicians: 12 women, 5 men).

Materials and procedure. The material and procedure were identical to Experiment 1, except that participants were now asked to read aloud the printed word, while ignoring the context. They were encouraged to respond as quickly and accurately as possible. The RTs were recorded by a voice key. During the session, the experimenter noted the error responses and the voice key dysfunctions.

Results and Discussion

Voice key dysfunctions led to exclude 2.82% of the data. Reading errors (0.02%) and RTs beyond three standard deviations of the mean (0.70%) were removed. Mean RTs for correct responses are shown in Figure 3.

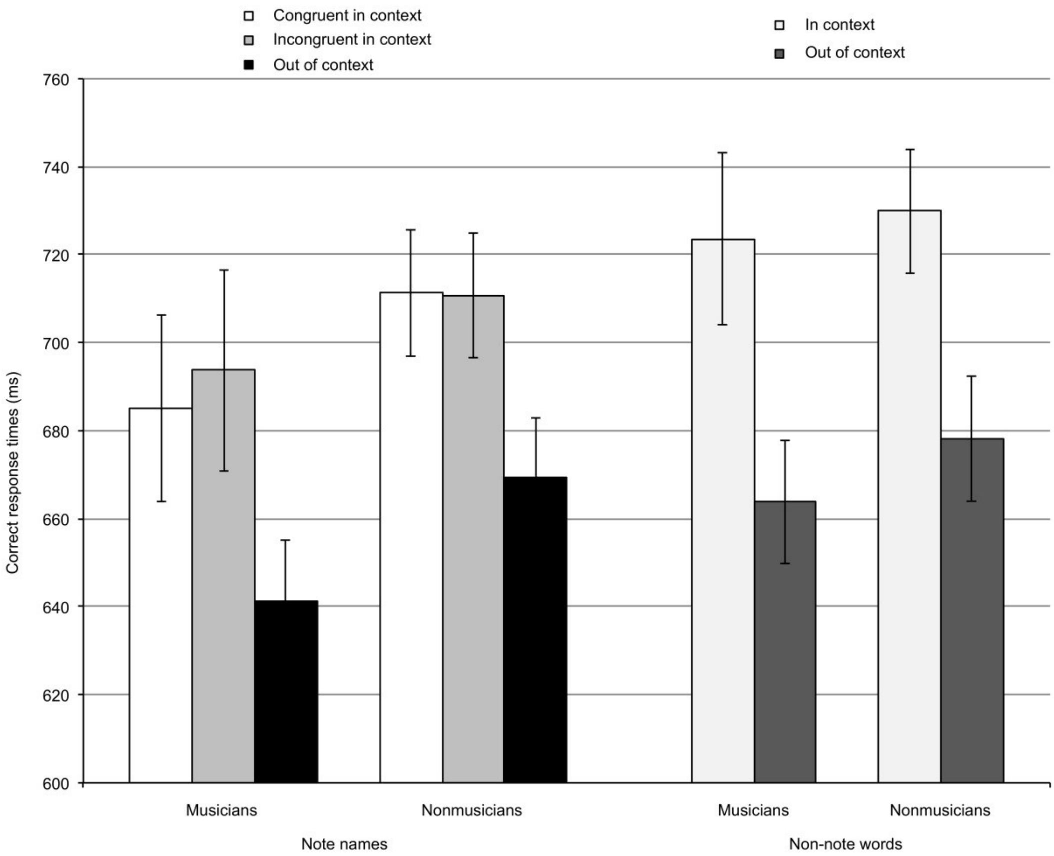

Figure 3. Correct response times as a function of Context and Musical Expertise in Experiment 2. Error bars indicate standard errors.

RTs to note names

For musicians, RTs were again significantly longer in incongruent- than in congruent-context condition, with a mean difference of 8.69 ms (SD = 16.46), $t(16) = 2.18$, $p = .045$, $d = .53$. There was no congruity effect for nonmusicians, $t(16) = 0.25$, $p = .80$. An ANOVA performed with Congruity (congruent-context, incongruent-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable revealed no main effect of musical expertise, $F(1, 32) = 0.75$, $p = .394$, and the congruity effect was only marginally significant, $F(1, 32) = 3.29$, $p = .079$, $\eta_p^2 = .093$. More importantly, there was a significant Congruity × Musical Expertise interaction, $F(1, 32) = 4.20$, $p = .049$, $\eta_p^2 = .116$. To sum up, we obtained clear evidence for a musical Stroop effect, with a congruity effect limited to musicians, hence replicating the results from Experiment 1.

Figure 3 also illustrates that reading note names was much faster in the out-of-context condition with regard to the congruent and the incongruent conditions for both musicians and nonmusicians, as in the go/no-go task used in Experiment 1. An ANOVA with Condition (congruent-context, incongruent-context, out-of-context) as a within-subject variable and

Musical Expertise (musicians, nonmusicians) as a between-subject variable gave only a main effect of condition, $F(2, 64) = 60.88$, $p < .001$, $\eta_p^2 = .655$. Neither musical expertise effect, $F(1, 32) = 1.10$, $p = .302$, nor the Condition × Musical Expertise interaction, $F(1,32) = 0.76$, $p = .472$, reached significance. Planned comparisons showed that RTs for the out-of-context condition were lower than both the incongruent, $F(1, 32) = 62.98$, $p < .001$, $\eta_p^2 = .663$, and the congruent conditions $F(1, 32) = 69.11$, $p < .001$, $\eta_p^2 = .684$.

RTs to non-note words

Crucially, there was also a clear effect of context when non-note words were used instead of note names in the two additional conditions. An ANOVA was performed on the RTs for non-note words with Context (in-context, out-of-context) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable. There was a significant and very large effect of context, $F(1, 32) = 180.50$, $p < .001$, $\eta_p^2 = .849$, whereas neither the musical expertise effect, $F(1, 32) = 0.24$, $p = .625$, nor the Context × Musical Expertise interaction, $F(1, 32) = 0.93$, $p = .341$, was significant. These results support our account of the results from Experiment 1 in terms of perceptual complexity.

Facilitation or interference?

As in Experiment 1, it remains possible to use the difference between musicians and nonmusicians in the out-of-context condition as a baseline for interpreting the differences between musicians and nonmusicians in the congruent- and incongruent-context conditions. In the out-of-context condition, musicians were 27.78 ms faster than nonmusicians. The difference for the congruent-context condition was nearly identical, $M = 26.27$, $F(1, 32) = 0.02$, $p = .885$. By contrast, the difference for the incongruent-context condition was numerically reduced ($M = 17.05$), although the reduction did not reach significance, $F(1, 32) = 0.82$, $p = .371$. This pattern is qualitatively similar to the pattern observed in Experiment 1, suggesting that congruence would elicit no beneficial effect in musicians, whereas incongruence would elicit a detrimental effect.

Finally, although the RTs of musicians were numerically shorter than the RTs of nonmusicians in all conditions, the analyses above show that the effect never reached significance. This suggests that the very strong effect observed in Experiment 1 was primarily due to the go/no-go task, in which the categorical membership decision was presumably very sensitive to the level of accessibility of the set of note names in the two groups of participants.

General Discussion

Musicians processed note names written inside note pictures in incongruent positions on a staff significantly slower in both a go/no-go task (Experiment 1) and a verbal task (Experiment 2) with regard to a condition in which note names were printed in congruent positions. Our results also showed that reading times of note names printed outside a staff cannot serve as an appropriate baseline to assess the proportion of the effect due to interference and facilitation. Indeed, inserting the note inside a staff slowed down reading times in Experiment 2, for both note names and non-note words, and in both musicians and nonmusicians, hence suggesting that the staff increased the perceptual difficulty of the task independently of its informational value for musicians. Assuming that the effect of perceptual complexity was identical for musicians and nonmusicians, however, our data suggest that congruence elicited no beneficial effect in musicians, whereas incongruence elicited a detrimental effect. This overall prevalence of interference over facilitation has been reported in nearly all other versions of the Stroop effect (e.g., MacLeod, 1991).

This study therefore gives the first compelling evidence for a musical Stroop effect, defined as a Stroop effect due to the interference generated by the automaticity of note processing in musicians. By trading word reading for note naming, the musical Stroop paradigm provides much greater flexibility in manipulating practice. Of particular interest is the fact that practice level can be decoupled from age and reading skill abilities, hence offering to researchers the possibility of manipulating a host of new variables in further studies. However, to be a useful tool of investigation, and not only a funny attraction, the musical Stroop effect (1) must fulfill a few additional criteria and (2), must have specific advantages with regard to other, non-standard Stroop-like procedures for research on automatisms. These two points will be discussed in turn.

Is the musical Stroop effect a manageable tool of investigation?

Using the musical Stroop effect requires that the potential study population is large enough. Musicians meeting the usual criteria of expertise for experimental studies on music cognition (in terms of years of practice in music school) are, admittedly, a small fraction of the general population. For instance in France, this fraction is estimated to about 2%. Given that the advantages linked to the musical Stroop effect (e.g., the possibility of a control group) imply to focus on abilities owned by a selective part of the general population, the only relevant question is: Although considerably reduced in percentage, is the remaining sample large enough for any practical purpose? Even a cursory estimation leads to a positive response. For the sake of illustration: A university comprising 20,000 students should include about 400 potential participants, a pool that is incomparably larger than the pool of patients typically available for neuropsychological investigations.

A second prerequisite is that this selective pool of participants does not differ from the general population along dimensions other than the amount of musical training. In particular, musicians could be more sensitive to interference than nonmusicians irrespective of the source of incongruency, leading to artificially inflate the differences between both groups. This specific argument does not hold, however. Existing evidence suggests that overall musicians are *less* sensitive to interference and have better executive control compared to nonmusicians (e.g., Bialystok & DePape, 2009; Travis, Harung, & Lagrosen, 2011). Although the limitations of the quasi-experimental distinction between the natural groups of musicians and nonmusicians make it difficult to strictly rule out the possibility of confounded variables (as in any other Stroop-like paradigms exploiting abilities acquired in real-world settings), our results support the assumption of interference due to extensive musical training.

A third prerequisite for an easy exploitation of the musical Stroop effect is that this effect is sufficiently robust and easily reproducible. The failure of Zakay and Glicksohn (1985) to observe a reliable effect in conditions conceptually similar to those of Experiment 2 (see Introduction) could suggest that the musical Stroop effect is an elusive phenomenon. We believe that this inference would be unwarranted, however. A part of the explanation for Zakay and Glicksohn's failure to get significant effects may stem from the fact that nonsignificance is inferred from a Scheffé's procedure, which is known to be highly conservative. What would have been the results of planned comparisons remains a matter of speculation. Irrespective of the method of analysis, it remains that Zakay and Glicksohn's study lacked of power, with a single response time per condition (i.e., for a homogeneous sequence of ten notes) for each participant. In addition, it is also worth noting that Zakay and Glicksohn's study did not meet the current methodological standards in Stroop research.

Among several other points,[8] all the notes were displayed as a continuous sequence on a sheet of paper. This latter feature is typical of early Stroop studies, in which RTs were measured with a manual chronometer for a whole sequence of stimuli belonging to the same condition. All recent experimental studies (including the present one) used a computerized item-by-item mode of presentation with mixed conditions, a procedure that minimizes the use of explicit strategies (MacLeod & MacDonald, 2000) and allows, among other advantages, to remove RTs for errors from RT analyses

Meeting current methodological standards, we observed a reliable musical Stroop effect in two independent experiments using different procedures. Although this effect was numerically small (within a 8-10 ms range), Cohen's *d*s indicated moderate (Experiment 2) to large (Experiment 1) effect sizes according to the conventional benchmarks. In addition, the numerical size of the effect may have been lowered by several factors. One of them is the proportion of congruent trials. In the standard color-word version, it has been shown that interference increases with the proportion of congruent trials (Lowe & Mitterer, 1982). In our experiments, the proportion of congruent trials was small (20%) due to the introduction of various control conditions (out-of-context notes and non-note words). Our results suggest that introducing these additional conditions in future investigations exploiting the musical Stroop effect does not look as mandatory, and we conjecture that a procedure involving only the two main experimental conditions, with half of the trials being incongruent and the other half congruent, would result in larger effects.

Arguably, the effect size would be also increased if a smaller sample of note positions was involved. In the present procedure, each note could be located on two positions on the staff in the congruent condition, hence exploring the knowledge of two full octaves. Using, say, only a few locations surrounding the treble key line would certainly tap the most automatic associations between note names and note positions.

Comparing the musical Stroop effect with other Stroop-like paradigms

In the introductory section, we have presented the advantages of the musical Stroop task for studying automatisms by contrast with the standard color-word or picture-word

[8] In particular, there was no counterbalancing or randomization (except for task order), which may have introduced noise. The presentation of the stimuli, such as shown in Zakay and Glicksohn's (1985, p. 417), Figure 1 is also questionable. The locations of the names of the notes were somewhat imprecise, with some overlaps. In addition, the note names were not written inside a note picture as in our materials, but simply written at note locations, which might have down played the effect.

Stroop tasks. Although these standard tasks are prevailing in the literature, there are a number of other variants, which have to be examined.

A first category of tasks follows the so-called "reverse Stroop" paradigm, in which the roles of word reading and color naming (or picture naming) are reversed. The stimuli are unchanged, but subjects are instructed to read the words, and any evidence of interference ought to be attributed to the automaticity of the processes engaged to deal with the irrelevant dimension. Since Stroop (1935), the Reversed Stroop Effect (RSE) has been known as an elusive phenomenon. By and large, an RSE has been obtained with oral responses if the usual conditions of reading were strongly degraded (e.g., Dunbar & MacLeod, 1984), or if participants were overtrained to inhibit the word prior to the experimental session (e.g., Dulaney & Rogers, 1994). Note that a reliable RSE has been obtained when verbal responses are replaced by responses made via word-labeled buttons (e.g., Blais & Besner, 2006, 2007). Irrespective of the size and robustness of the RSE, however, abilities such as color naming and picture naming are certainly not good candidates for studying the formation of automatisms, given the difficulty of controlling what counts as "practice" for these tasks in real-world settings.

Other paradigms do not involve reading at all. For instance in studies involving numerosity, counting is impaired whenever the stimuli being counted are incongruent numerals (e.g., Heine et al., 2010; Shor, 1971). But manipulating practice with number is hardly easier than manipulating reading practice, because everyone acquires both abilities at school within the same age span. Other studies have attempted to circumvent the problem by using arbitrarily structured tasks in which training is entirely performed in the laboratory context (e.g., Regan, 1981). This indeed allows a perfect control over practice, but at the expense of losing what makes the Stroop task so attractive, namely its propensity to assess the consequences of very extensive practice over years.

A last to-be-examined paradigm seems, at first glance, more promising. Several studies have used a standard Stroop paradigm except that words belonged to a second language. Second language learning often occurs later than reading acquisition, involves extensive practice, and arguably, the amount of practice can be controlled as well as the learning of music. We do not intend to deny the interest of relying on the learning of a second language to research on automatisms, as successfully performed in a few earlier studies (e.g., Tzelgov, Henik, & Leiser, 1990). However, it is worth noting that the bulk of the Stroop literature involving a second language is devised to enlighten various issues on bilingualism, rather than to improve our understanding of automatization. Of course, as such, this does not mean that the paradigm would be unsuitable to studying automatization, insofar as the current focus on bilingualism could only reflect researchers' main interest. However, the fact that many issues on bilingualism are currently unsolved is actually damaging for drawing clear conclusions on automatisms. To illustrate: What is assumed to be "automatized" with the practice on a second language? Is it the link between the words of each language (e.g., RED-rojo, with rojo activating RED and RED activating an interfering response)? Is it the direct access to the concepts (with rojo directly activating an interfering response)? A response to these questions depends on theoretical options (see especially the "word association" model vs. the "concept mediation" model of bilingual memory, e.g., Kroll & Stewart, 1994). At a more elementary level, what is the role of learning the grapheme/phoneme mapping? This role is presumably null whenever the languages share most of their alphabet (e.g., English/Spanish), but the response is less obvious when the two languages involve different symbols, as English and Chinese (see for instance Chen & Ho, 1986). Presumably, the weight of all these factors changes throughout practice, hence making the study of a putative "automatization" process all the more complex. In any case, it would be a huge oversimplification to consider that using a Stroop paradigm with a second language allows reproducing what occurs initially with one's first language, with the advantage of decoupling age and level of practice.

A final comment about reading

Although we focused throughout this paper on the automatism of note naming, it is worth adding that our study is endowed with interesting implications with regard to the automatism of reading. As mentioned above, earlier studies having explored whether reading could be impaired by an incongruent context (i.e., the reverse Stroop paradigm) have mostly concluded in a negative way. This is generally interpreted as evidence that "the extreme automaticity of word reading is very difficult to overcome" (MacLeod, 1998, p. 207). In this context, our positive results appear as an exception. Reading times were slowed down by incongruent note positions, even though note names were clearly printed, in a font that certainly surpassed in size the fonts used in most books or newspapers, and without pretraining to inhibit word processing. This suggests that the extreme automaticity of word reading does not protect it against interference, provided that the competing process is itself strongly automatic.

The intriguing possibility that automaticity of behavior does not make it immune to interference is strengthened by the results from a recent paper by Akiva-Kabiri and Henik (2012). In this paper, musician participants, among other tasks, were asked to read the name of notes while hearing a tone that could either correspond to the note (congruent condition) or not (incongruent condition). The most relevant part of the results for the present concern was that absolute pitch possessors showed a significant congruency effect. This effect demonstrates that pitch identification in absolute pitch possessors (a mainly inherited ability) was impossible to suppress, thereby interfering with word reading.

Interestingly, in the Akiva-Kabiri and Henik's (2012) experiment, absolute pitch possessors were unaffected by the written note name when they were asked to label the auditory tone, as if the congruency effect could act only with a single dimension at a time. Whether such an effect is also obtained with the musical Stroop paradigm, in which the automaticity of note naming is due to extensive practice, remains to be investigated. More generally, the possibility of putting two automatisms against each other should open a new way of investigation for further research.

References

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical stroop effect in absolute pitch possessors. *Experimental Psychology, 59*, 272-278.

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 565-574.

Blais, C., & Besner, D. (2006). Reverse stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1345-1353.

Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review, 14*, 466-469.

Catling, J. C., Dent, K., Johnston, R. A., & Balding, R. (2010). Age of acquisition, word frequency, and picture-word interference. *The Quarterly Journal of Experimental Psychology, 63*, 1304-1317.

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 397-401.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 470-484.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 622-639.

Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics, 65*, 367-378.

Heine, A., Tamm, S., De Smedt, B., Schneider, M., Thaler, V., Torbeyns, J., et al. (2010). The numerical stroop effect in primary school children: A comparison of low, normal, and high achievers. *Child Neuropsychology, 16*, 461-477.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 39*, 149-174.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided Attention in a Stroop task. *Canadian Journal of Psychology, 36*, 684-700.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7*, 485-495.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*, 201-211.

MacLeod, C. M., & MacDonald, P. A. (2000). Inter-dimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383-391.

Nealis, P. M. (1973). The Stroop phenomenon: Some critical tests of the response competition hypothesis. *Perceptual and Motor Skills, 37*, 147-153.

Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology, 29*, 201-209.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology, 54*, 251-282.

Regan, J. E. (1981). Automaticity and learning: Effects of familiarity on naming letters. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 180-195.

Schulz, T. (1979). Components of the reaction time Stroop-task. *Psychological Research, 40*, 377-395.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing : II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology, 93*, 281-288.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109*, 373-392.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a biligual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 760-771.

Travis, F., Harung, H. S., & Lagrosen, Y. (2011). Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians:

Interpreted in light of a Unified Theory of Performance. *Consciousness and Cognition, 20*, 1256-1264.

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology, 39*, 414-423.

Réponse aux commentaires

Is the Musical Stroop Effect able to keep its promises?
A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014)

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
CNRS-UMR 5022, University of Bourgogne, Dijon, France

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014). Is the Musical Stroop Effect able to keep its promises? A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology, 61*, 80–83.

Corresponding author:
Laurent Grégoire
Université de Bourgogne
LEAD-CNRS UMR5022
Pôle AAFE, Esplanade Erasme
21000 Dijon, France

E-mail: laurent.gregoire@u-bourgogne.fr

Résumé

L'étude réalisée par Grégoire, Perruchet, et Poulin-Charronnat (2013) a suscité des commentaires de la part de plusieurs chercheurs: Akiva-Kabiri et Henik (2014; annexe 1), Gast (2014; annexe 2), Moeller et Frings (2014; annexe 3), et Zakay (2014; annexe 4). La plupart des critiques présentées dans les quatre commentaires sont basées sur une incompréhension de notre procédure, ou des postulats discutables. Nous maintenons que l'effet Stroop musical offre des possibilités prometteuses pour la recherche sur les automatismes, tout en reconnaissant que la procédure actuelle ne permet pas de distinguer aisément la facilitation et l'interférence.

Abstract

Grégoire, Perruchet, and Poulin-Charronnat (2013) claimed that the Musical Stroop task, which reveals the automaticity of note naming in musician experts, provides a new tool for studying the development of automatisms through extensive training in natural settings. Many of the criticisms presented in the four commentaries published in this issue appear to be based on a misunderstanding of our procedure, or questionable postulates. We maintain that the Musical Stroop Effect offers promising possibilities for further research on automaticity, with the main proviso that the current procedure makes it difficult to tease apart facilitation and interference.

Introdution

The basic arrangement of the experiments reported in Grégoire, Perruchet, and Poulin-Charronnat (2013) comprised a staff with a note in various positions. A name of a note was printed inside the note. In the congruent condition, the note name was congruent with the note position, whereas in the incongruent condition, name and position were incongruent. The main result was that musicians asked to read aloud the written name of the notes showed impaired processing in the incongruent condition with regard to the congruent condition, a result coined as the *Musical Stroop Effect* (MSE). In this reply, we focus on the points that have been presented as potential limitations of the procedure in the four previous commentaries. We thank the commentators for other helpful and constructive comments, which space limitation does not allow discussing.

The musical Stroop task is just another Stroop-like test

Zakay (2014) writes that the musical Stroop task "is no more than another Stroop-like test". The material is indeed similar to a standard picture-word interference task, as repeatedly noted by **Moeller and Frings** (2014; **M&F**), but, crucially, the MSE is the *reverse* of the effect explored in most Stroop-like tests. Reading is involved, but as the object, rather than the source, of interference. The procedure was designed to investigate the automaticity of note naming in musicians, in the same way as the classical Stroop task investigates the automaticity of word reading. We agree with **Zakay** that the musical Stroop task does not replace the standard color-word version, but this was not our objective. More modestly, we intended to provide a tool that would be better suited than the other Stroop tasks to address a specific question in future research: How does Stroop interference evolve with practice? The key point is that the emergence of automaticity in musicians provides the opportunity of better control on the conditions of training than reading acquisition, notably because the level of musical expertise can be easily decoupled from age and academic level. As **M&F** rightly point out, there are several conflicting hypotheses about how interference may evolve with increasing expertise, and exploring this evolution looks as promising.

There is no note-picture note-name association

Zakay, and to a lesser extent **M&F,** contend that the note name would not be activated by a note picture, thus negating the very existence of the automatism we were tracking in our experiments. A major problem with this contention is that, if right, it would leave unexplained why an MSE occurred in our experiments.

M&F invoke as a definitive argument for justifying that the name of the note cannot be activated: "Just think of a trumpeter!" It must be understood that assuming the automaticity of note naming in musicians does not amount to claim that a note picture generates an irrepressible need to name aloud the note, which would be obviously incompatible with the practice of a wind instrument, among other activities. If an overt response were required, then the classical Stroop effect would not exist in the first place, given that reading aloud is certainly infrequent in adults. As suggested by **Gast** (2014), it is even possible that the note name does not really evoke a *response*, even subvocal. The MSE would come from the learned association of note and note name, conceived as a stimulus-stimulus relationship, which would generate interference during the encoding phase when the note/note name contingency is broken. This is a sensible hypothesis, which warrants further investigations.

Of course, we are not asserting that the mastery of note naming is required for any form of musical performance. As **Akiva-Kabiri and Henik (2014; A&H)** observe, note labeling may not be needed for playing by ear, or still playing from memory or improvising. Our contention is, however, that note naming is a key component of music reading. As Hodges and Nolker (2011) wrote: "Although there are many oral musical traditions and practices, music reading hold a special place in contemporary music education curricula. [...] Virtually all beginning instrumental method books and private instructional books [...] have sections on music reading, as do general music basal series" (p. 61).

Trading verbal responses against manual responses would be better

Starting from the postulate that musical training is primarily directed toward the automatization of the motor programs involved in the practice of a musical instrument, several commentators suggest requiring the production of the note on a musical instrument, as a complement **(Gast)** or as a better alternative (**M&F, Zakay)** to verbal responses. Although appealing at first glance, this suggestion overlooks the fact that the stimulus to which participants are asked to respond in the musical Stroop task is *not* a note on a staff, but a written note name. To illustrate what happens in these conditions, let us consider Zakay and Glicksohn's (1985) study. Among the many conditions of their experiment with a sample of pianists, the authors introduced verbal (reading aloud) and motor (pressing the appropriate key on the piano) modes of responding to written note names. It turned out that motor responses were, on average, considerably longer than verbal responses (650 ms vs. 430 ms, respectively). The reason is straightforward: A written note name does not provide sufficient information to trigger a unique motor response. In particular, a note name does not specify the octave, hence obliging pianists to an arbitrary choice between seven or eight possible piano keys for each note. It looks unlikely that response production processes that imply time-consuming intentional decisions could be impacted by motor automatisms.

In apparent contradiction with our analysis, however, the data reported with motor responses in Zakay and Glicksohn's Table 1 showed a difference between incongruent and congruent items, which even largely exceeded in amplitude the standard Stroop effect (287 ms). Although the authors made neither descriptive nor inferential analyses of this effect, **Zakay** retrospectively describes it as an earlier demonstration of an MSE with a motor response. This claim is questionable. An alternative, and much more plausible explanation for the 287 ms difference appears when looking at the materials used by Zakay and Glicksohn, which is partially reproduced in Figure 1 (see also our discussion of this study in Grégoire et al., 2013). The crucial point is that the time-consuming selection of a specific key is no longer necessary in the congruent condition because, by contrast with the incongruent condition, the note location on the staff now designates a correct and unique key for the response. Maybe another procedure would be successful in revealing an MSE with motor responses, but the design remains to be invented. Our current feeling is that the lack of a one-to-one mapping between a note name and a motor response (whatever the musical instrument) raises an insurmountable obstacle to simply trade verbal mode against motor mode of responding in the musical Stroop task.

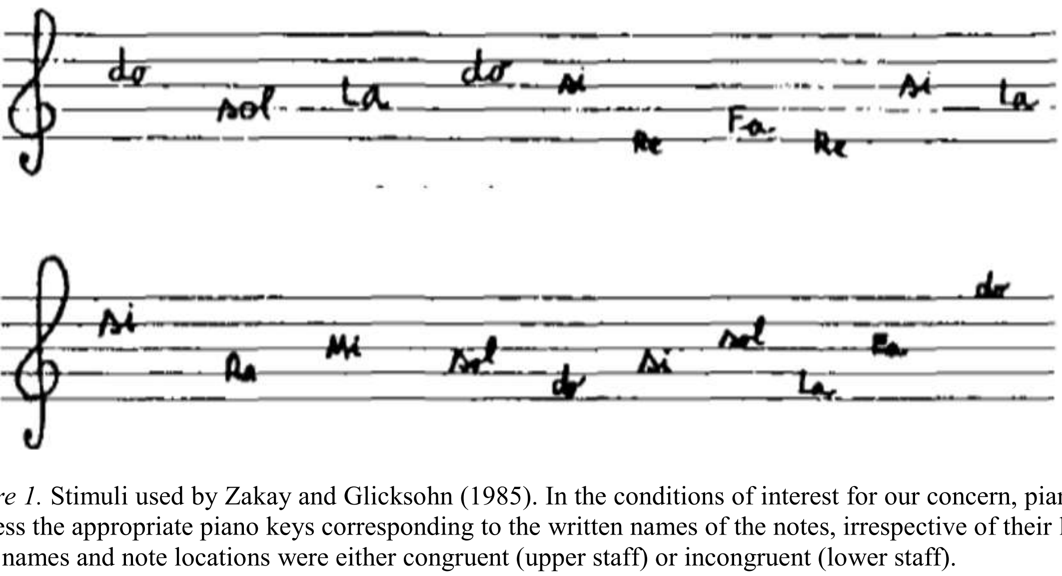

Figure 1. Stimuli used by Zakay and Glicksohn (1985). In the conditions of interest for our concern, pianists had to press the appropriate piano keys corresponding to the written names of the notes, irrespective of their location. Note names and note locations were either congruent (upper staff) or incongruent (lower staff).

The two competing dimensions are integrated (or separated)

According to **M&F**, the note picture and the written note name that is written inside belong to the same object. **M&F** worry: "Thus, it cannot be ruled out that the interference of note naming on note name reading would diminish if written note name and note would not belong to the same object. This seems to be the case for other variants of the Stroop effect and would argue against a strong claim of automatic processing." **Zakay** describes our arrangement in a diametrically opposite way on the separated/integrated dimension, thus asserting that our dimensions are separated. The point is that, ironically, even though the diagnoses of **M&F** and **Zakay** are opposite, their conclusions are identical: Our design is defective.

These criticisms are puzzling because, to our knowledge, no one has ever questioned the validity of the conclusions stemming from a Stroop study on the ground that the level of integration of the two competing dimensions was too strong (**M&F**) or too weak (**Zakay**). It has been well-documented that integrated dimensions produce the largest amount of Stroop interference, and that interference decreases when the competing stimuli move farther one from each other (MacLeod, 1991). Therefore, it is highly predictable that if the picture shown in the right panel of **M&F**'s Figure 1 were used instead of the currently used picture, the MSE would be reduced. This indeed could be put forth against a strong claim about the automaticity of note naming, as commonly acknowledged for reading since the 1980's, but we fail to see how this would impact our own work. To be sure, we have never claimed that the automaticity of note naming in musicians would be stronger than the automaticity of reading.

Teasing apart facilitation and interference

A&H point out the difficulty to tease apart interfering and facilitating components of the MSE. The question is: Where a note name should be printed to give a baseline reading time to which performance in congruent and incongruent conditions could be compared? A seemingly obvious response is "out of the staff", but our results revealed that reading a printed word out of the staff is much shorter than reading the same word inside a staff, even for nonmusicians, presumably due to the perceptual complexity induced by the staff.

A&H recommend printing the note name on an inverted staff. This would indeed control for some physical components of perceptual complexity, and we are relatively confident that this condition would equalize the reading times between the neutral, congruent, and incongruent conditions for nonmusicians. We are more skeptical about the neutrality of an inverted staff for musicians, because we suspect that musicians may either neglect the reversal (hence coding the note as in the congruent and incongruent trials) or take the reversal into account, hence processing the stimuli as if the whole score was displayed upside down.

As a result, we agree that the difficulty of dissociating facilitation and interference in the MSE is an important limitation of the paradigm, and we acknowledge that our almost exclusive focus on the interference component is a questionable shortcut. This being so, in the same way as we see no reason to deny that some facilitation could account for a part of the MSE as in other Stroop paradigms, we see no reason to believe that this part would be larger than in other paradigms. MacLeod (1991) concludes from his review on this issue that "facilitation [in the congruent condition] is much less than the corresponding interference in the incongruent condition" (p. 175). He adds: "and the choice of control condition may be crucial". The last sentence usefully recalls that even if the choice of a control condition may be especially problematic for the MSE, it would be wrong to believe that other Stroop paradigms are totally free from similar intricacies. Some investigators have even questioned the use of a baseline to measure facilitation and interference effects in the standard color-word task (e.g., Lindsay & Jacoby, 1994).

Does music practice offer the best opportunity to track automatism formation?

As mentioned above, the main objective of replacing reading by note naming in the Stroop task is the possibility of better control on the conditions of training. Several commentators note various problems that could arise when trying to control the level of practice in music, nevertheless. For instance, **A&H** note the variability of the musical notation systems across cultures and instruments. Along the same line, **Gast** points out that the learning of musical notation is not a linear and incremental process, which may complicate the interpretation of learning curves across years of musical school. These are useful caveats. Using the number of years of practice as the single criterion for selecting participants is certainly insufficient, and further studies should refine the selection criteria. Measuring the speed and accuracy of note naming with a standard test could be a useful complement.

A&H suggest that investigating the automaticity of reading during second language acquisition could provide a better control over training than the MSE. Our feeling is that the practical feasibility of one or another paradigm may depend on conditions specific to each country. With regard to the current landscape in France, the acquisition of second language is hard to decouple from age and academic level, whereas the organization of musical teaching ensures nice conditions of investigation. Indeed, musical teaching is mainly provided by music schools that recruit people of any age, and which place great emphasis on music reading. In addition, Grégoire et al. (2013) noted that many conceptual issues on bilingualism are currently unsolved, and that the complexity of the involved processes could be damaging for drawing clear conclusions on automatisms. That being said, we fully share the view that further research following this approach would be worthwhile.

Gast usefully recalls that creating an automatism does not necessarily requires years of practice, and that some Stroop-like effects can be observed after a quite limited amount of training in laboratory conditions. As **Gast** notes, this approach has the obvious advantage of ensuring the best experimental control. It remains to be seen whether the very same phenomena are observed at a micro-level in a one-session laboratory task and at a macro-level after years of consistent practice. Some Stroop-like effects have been observed with arbitrary mapping in laboratory, as **Gast** mentions, but to our knowledge, newly acquired automatisms remain unable to interfere with reading (MacLeod, 1998), thus suggesting that laboratory practice cannot serve as a substitute for extensive practice in real-world conditions. For instance, **Gast** rightly asserts that "MacLeod and Dunbar (1988) showed that newly learned naming responses can lead to interference in a Stroop task after five one-hour training sessions", but it is important to add that the observed pattern of interference did *not* imply word reading. Again, we believe that conceiving the different paradigms in terms of complementarity is a better approach than reasoning in terms of competition and exclusive alternatives.

Conclusion

Zakay worries that "when names of notes are written within note pictures on a staff, a very particular and unusual non-ecological condition is created". We agree. But, except for experimental psychologists, is it more usual to be asked to name the color of incongruent color words? The grounding principle of a Stroop task is to involve an unusual situation, because this condition is necessary to mislead our automatisms. Of course, there are other paradigms for assessing automaticity, as **Gast** recalls. However, insofar as one includes Stroop paradigms among the worthwhile approaches of automaticity, we did not find in the commentaries any arguments that could invalidate the ability of our procedure to fulfill its primary objective, namely providing a better control on the conditions and the level of training than most classical versions.

References

Akiva-Kabiri, L., & Henik, A. (2014). Additional insights (to-be-completed).

Gast, A. (2014). What is learned, and when? (to-be-completed).

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The Musical Stroop Effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Hodges, D., & Nolker, B. (2011). The acquisition of music reading skills. In R. Colwell & P. Webster (Eds.). *MENC Handbook of Research on Music Learning, Volume II: Applications* (pp. 61–91). Oxford: Oxford University Press.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 219–234.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*, 201–211.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126–135.

Moeller, B., & Frings, C. (2014). How automatic is the Musical Stroop Effect? (to-be-completed).

Zakay, D. (2014). Can the "musical stroop" task replace the classical Stroop task? (to-be-completed).

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology, 39*, 414–423.

Etude 2

About the Unidirectionality of Interference: Insight From the Musical Stroop Effect

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
CNRS-UMR 5022, University of Bourgogne, Dijon, France

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (in press). About the Unidirectionality of Interference: Insight From the Musical Stroop Effect. *Quarterly Journal of Experimental Psychology.*

Author Note

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat, Université de Bourgogne, LEAD/CNRS, UMR5022, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France.

This research was supported by the grant ANR-09-JCJC-0129 and the grant Conseil Régional de Bourgogne N° 2012-9201AAO05S01353.

Correspondence concerning this article should be addressed to Laurent Grégoire, LEAD-CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 2100 Dijon, France. E-mail: laurent.gregoire@u-bourgogne.fr

Résumé

Le paradigme de Stroop musical, parmi les nouvelles possibilités qu'il offre à l'exploration expérimentale, permet de confronter deux traitements hautement automatisés après une pratique intensive et naturelle : la lecture de mots et la dénomination de notes.
La littérature mentionne abondamment l'asymétrie de l'effet Stroop, selon laquelle les mots interfèrent sur la dénomination de la couleur (ou la dénomination d'images) alors que la couleur (ou les images) n'altère(nt) pas la lecture des mots. Plusieurs études impliquant un paradigme de type Stroop suggèrent qu'un effet Stroop inversé pourrait être observé, mais au détriment de l'effet Stroop lui-même, comme si l'interférence était intrinsèquement unidirectionnelle. Elle résulterait de l'altération du processus le moins automatisé par le processus le plus automatisé, ce qui est consistant avec le modèle de Cohen, Dunbar, et McClelland (1990). Néanmoins, les résultats que nous obtenons dans cette étude mettent en évidence un effet Stroop bidirectionnel. Nous avons exposé des musiciens aux situations congruentes et incongruentes du paradigme de Stroop musical en leur demandant, soit de lire le mot en ignorant la position de la note sur la portée, soit de dénommer la note sans tenir compte du mot écrit à l'intérieur. La grande majorité des participants (24 sur 26) ont manifesté un effet de la congruence dans les deux tâches. A première vue, ces résultats vont à l'encontre du modèle de Cohen et al. (1990), mais des analyses complémentaires ont rapporté qu'ils étaient cohérents avec un des principes clés de ce modèle, selon lequel le patron d'interférence dépend de la force relative des deux traitements en compétition. Nous suggérons que la différence entre les résultats collectés avec le paradigme de Stroop couleur-mot (ou image-mot) et ceux obtenus avec notre paradigme note-mot chez les musiciens révèle l'importance de la contingence stimulus-réponse dans la formation des automatismes et leur pouvoir interférant.

Abstract

The asymmetry of interference in a Stroop task usually refers to the well-documented result that incongruent color words slow color naming (Stroop effect) but incongruent colors do not slow color word reading (no reverse Stroop effect). A few other studies have suggested that, more generally, a reverse Stroop effect can be occasionally observed but at the expense of the Stroop effect itself, as if interference was inherently unidirectional, from the stronger to the weaker of the two competing processes. We describe here a situation conducive to a pervasive mutual interference effect. Musicians were exposed to congruent and incongruent note name/note position patterns, and they were asked either to read the word while ignoring the location of the note within the staff, or to name the note while ignoring the note name written inside the note picture. Most of the participants exhibited interference in the two tasks. Overall, this result pattern runs against the still prevalent Cohen, Dunbar, and McClelland's (1990) model of the Stroop phenomenon. However, further analyses lend support to one of the key tenets of the model, namely that the pattern of interference depends on the relative strength of the two competing pathways. The reasons for the impressive differences between the results collected in the present study and in the standard color-word (or picture-word) paradigms are also examined. We suggest that these differences reveal the importance of stimulus-response contingency in the formation of automatisms.

Introduction

Grégoire, Perruchet, and Poulin-Charronnat (2013) devised a new version of a musical Stroop task (for an earlier musical Stroop paradigm investigating motor automatisms in pianists, see Stewart, Walsh, & Frith, 2004). The basic arrangement comprises a staff with a note in various positions (see Figure 2). A name of a note is printed inside the note. For the congruent condition (Figure 2a), the note name is congruent with the note position on the staff, whereas in the incongruent condition (Figure 2b), note name and position are incongruent. Musicians asked to read the written names of the notes showed impaired processing in incongruent condition with regard to congruent condition. This *Musical Stroop Effect* (MSE) attests to the interference[9] generated by the automaticity of note naming in musicians.

The primary motivation for studying the automaticity of note naming, instead of the automaticity of word reading as do the conventional version of the Stroop paradigm and most Stroop-like tasks (e.g., Flowers, Warner, & Polansky, 1979; Glaser & Düngelhoff, 1984; Virzi & Egeth, 1985), stemmed from the greater flexibility in manipulating practice. Of particular interest is the fact that practice level can be decoupled from age and reading-skill abilities, hence offering to researchers the possibility of manipulating a host of new variables (Grégoire, Perruchet, & Poulin-Charronnat, 2014). However, other questions are linked to the fact that, in the MSE, reading is not put outside overall; its status is just reversed, from the status of the interfering process to the status of the interfered process. A Stroop effect in which reading is interfered by another process is often coined as a *Reverse Stroop Effect* (RSE).[10] In the remainder of this introduction, we first outline the main findings regarding the RSE, as well as their theoretical implications, then we examine why the Grégoire et al.'s musical Stroop is a unique test-bed to address one of the major issues raised by this literature.

The reverse Stroop effect

Stroop (1935) was the first to examine whether *reading* color words could be impaired

[9] For the sake of simplicity, except when specified otherwise, the difference between incongruent and congruent conditions will be coined here as interference, even though a part of the effect may arguably be due to facilitation in congruent condition.

[10] This terminology is conventional (e.g., MacLeod, 1991), but not without its own shortcomings, notably because it applies only to situations in which reading is one of the two competing processes. Given that an overwhelming proportion of the Stroop literature involves reading, however, this is not a strong limit, and this definition is used throughout this paper.

by an incongruent color-word combination. He reported an RSE, but only after extensive practice to name the color of incongruent words. In addition, this effect was short-lived, as evidenced by the disappearance of the RSE a few days after the participants stopped practicing. Dulaney and Rogers (1994) replicated this result, and showed that the RSE obtained after practicing the color-naming task for incongruent words was due to the development of a "reading suppression response". Participants seemingly did not automatize color naming but instead acquired a transitory capacity to inhibit the reading response. This conclusion was supported by Ellis, Woodley-Zanthos, Dulaney, and Palmer (1989), who observed that participants previously trained to name the color of control items (a series of colored Xs), a task that offers no opportunity to inhibit reading, gave no evidence for an RSE. More recently, MacLeod (1998) failed to observe an RSE with the now standard single-item procedure despite very extensive training to name incongruent color words. He attributed the fleeting effect reported in earlier studies to the use of the multiple-item version of the task, in which trial types are not mixed. To sum-up, the pervasive failure to get an RSE under usual reading conditions has led to construe the asymmetry of interference as a ubiquitous property of the reading processes.

The overall picture, however, is more complex. In fact, an RSE has been obtained in color-word Stroop versions when the usual conditions of reading were strongly degraded (e.g., Dunbar & MacLeod, 1984, Melara & Mounts, 1993), or still when verbal responses were replaced by motor responses (e.g., Blais & Besner, 2006, 2007, Durgin, 2003; Melara & Mounts, 1993). Moreover, other studies (e.g., Akiva-Kabiri & Henik, 2012; Palef & Olson, 1975) also observed an RSE in Stroop-like procedures that involved a reading task but another competing process than color naming. At first glance, these data could simply suggest that Stroop asymmetry is not as ubiquitous as once thought. But a closer scrutiny of the few studies investigating concurrently the RSE and the standard effect completes this conclusion with a more intriguing observation.

Let us consider the recent study of Akiva-Kabiri and Henik (2012). In this study, musician participants were asked to read the name of notes while hearing a tone that could either correspond to the note (congruent condition) or not (incongruent condition). Absolute pitch possessors (a mainly inherited ability) showed a significant RSE. Reading was impaired when the written word was incongruent with the tone. By contrast, those participants did not show interference when they were asked to label the auditory tone in presence of an incongruent written note name. Moreover, control musicians without absolute pitch showed the inverse pattern: They were not affected by an incongruent tone when reading a note name, but they were affected by an incongruent written note name when labeling the tone. In a nutshell, the results from Akiva-Kabiri and Henik suggest that, by and large, the Stroop effect and the RSE are mutually exclusive.

A few earlier studies suggested similar conclusions. Palef and Olson (1975) used a paradigm in which the words *above* and *below* were presented either above or below the fixation point. In their first experiment, reading the word showed interference from incongruent locations, hence evidencing an RSE. Strikingly, however, incongruent word meanings did not interfere with decisions about spatial position. The authors reasoned that this asymmetrical effect could be due to the fact that spatial position was processed faster than word meaning. To test this hypothesis, they designed a second experiment in which the relative duration required for each task was reversed. The words were now displayed at the fixation point, with an asterisk serving of reference to the spatial judgment being either above or below the words. Under these conditions, decisions about spatial position took longer than reading the words, and concurrently, the direction of interference was reversed. In other words, everything happened as if interference could act in either direction, but with a single dimension at a time as a function of conditions. Some amount of interference was bidirectional, however, but only in a subgroup of participants in Experiment 2 for whom, due to the chronological ordering of the tasks, the processing times turned out to be approximately equal between reading and judgments of spatial position.

A similar pattern of data emerged from Melara and Mounts's (1993) experiments with the standard color-word Stroop paradigm. When the readability of the color word was reduced (small fonts were read long away from the screen), the authors observed a large RSE on both response times and errors, but at the expense of the standard effect, which did not reach significance (Experiment 2). Thus the usual pattern of asymmetry was inverted. In two other experiments (Experiments 1 and 4), the authors equalized the perceptual discriminability of words and colors, again by shrinking the visual angle of the stimuli (but to a lesser extent than in Experiment 2). Under these very specific conditions, both the standard effect and its reverse were observed, but these effects were small and volatile. To borrow the authors' words: "subjects displayed meager, fleeting, and generally symmetrical effect of irrelevant variation" (p. 642).

The final picture is that the asymmetry of the Stroop effect cannot be understood as the ubiquitous prevalence of reading over any other process, even if it is its most usual expression. However, the asymmetry could be redefined as the fact that whatever the two competing processes, interference may go in either direction as a function of participants and experimental conditions, but, crucially, in a single direction at a time.[11] A few studies reported statistically significant bidirectional interference, but with effects that were small, ephemeral, and circumscribed to very specific experimental arrangements. As pointed by Cohen et al. (1990), this pattern of data, based on group statistics, may reflect a mixture of unidirectional effects going in opposite direction for different subjects, and therefore cannot be taken as a decisive argument to refute the principle of strict unidirectionality.

Theoretical implications

The Cohen et al.'s (1990) connectionist model is still acknowledged as one of the prevalent models of the Stroop effect (e.g., Blais & Besner, 2006, 2007; MacLeod & MacDonald, 2000; Protopapas, Archonti, & Skaloumbakas, 2007; Roelofs, 2005). Cohen et al. did not refer to Palef and Olson (1975), and they were obviously unaware of later studies (Akiva-Kabiri & Henik, 2012; Melara & Mounts, 1993). However, one of the main motivations for the development of their model was the report of related data by MacLeod and Dunbar (1988). Although MacLeod and Dunbar's paradigm did not involve reading, and hence does not concern the RSE as defined above, their results are highly relevant for our concern. In MacLeod and Dunbar, a specific color name was arbitrarily assigned to each of four black-and-white polygons, and participants were instructed to learn these names along successive practice sessions. After different amounts of training, shapes were presented in colors that were either congruent or incongruent with their name. Participants carried out two tasks: They had to name the shapes ignoring their colors, and to name the colors disregarding the shapes. Early in training, color had an effect on shape naming but not vice versa. After extensive training in shape naming, the opposite pattern of results was observed: Shape had an effect on color naming but not vice versa. Therefore, the effect and its reverse were mutually exclusive, at least when the extreme ends on the level of practice were considered (i.e., Day 1 and Day 20 in their Experiment 3).

How does Cohen et al.'s (1990) model account for MacLeod and Dunbar's (1988)

[11] Using two logographic scripts, Japanese and Chinese, Verdonschot, La Heij, and Schiller (2010) reported standard Stroop effects in a picture-word interference task. However, when investigating the RSE, they observed that reading Japanese kanji was *shortened* by the incongruent pictures with regard to unrelated pictures. No effect at all was observed with Chinese hànzì. Irrespective of the interpretation of these findings, they concur to generalize the asymmetry of the Stroop effect to nonalphabetic writing systems.

results, and notably for the fact that the direction of interference is reversed even though the strength of the color-naming pathway remains, presumably, constant? In this model, the attributes of automaticity depend on the strength of a processing pathway, which is itself a function of the amount of training received in the task at hand. The critical point is that when two pathways conflict, the resulting pattern of interference does not depend on the absolute strength of each pathway, but on the *relative* strength of the two pathways. To quote Cohen et al., it is "the relative strength, compared with a competing pathway [that is] important in determining whether a process will produce or be subject to interference in a Stroop-like task" (p. 348). Figure 1 illustrates how the direction of interference between two processes A and B may be reversed even though the strength of B does not vary (Situations 1 vs. 4 in the figure).

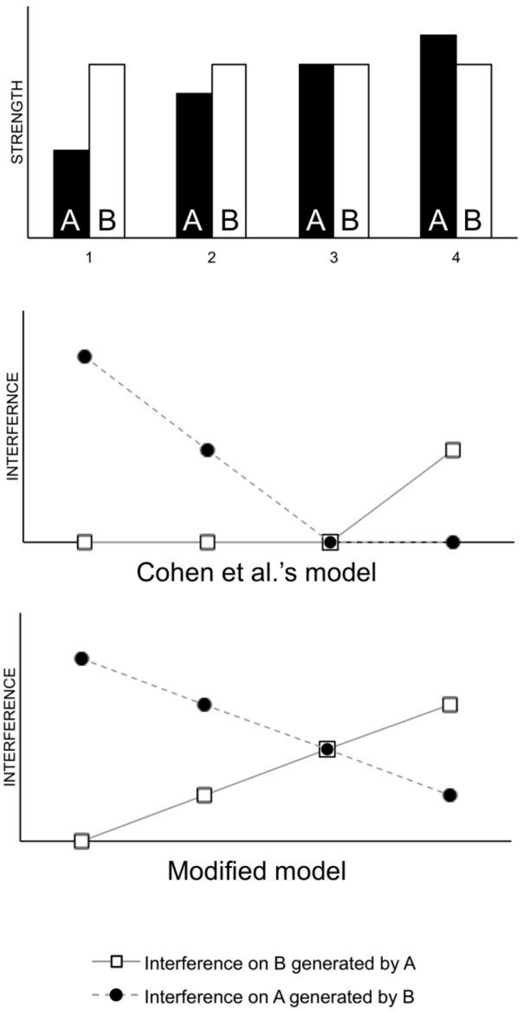

Figure 1. Schematic predictions of the Cohen et al.'s model and its modified version, in four situations, in which the strength of Process B is kept constant, while the strength of Process A gradually increases. In both models, the amount of interference (on the y-axis, arbitrary units) depends on the relative strength of the competing processes. However, in Cohen et al.'s model, only the stronger of the two processes generates interference, giving unidirectional interference when the strength of the two processes differs and no interference at all when the two processes have equal strength. The modified, "softened" model allows some amount of interference from the weaker process on the stronger process.

Nevertheless, there is some discrepancy between MacLeod and Dunbar's (1988) empirical data and the model's predictions, regarding what occurs when the two competing processes are of equal strength. In MacLeod and Dunbar, each dimension influenced the other to the same extent after an intermediate amount of practice (i.e., Day 5). In the model, there is no possibility for a process A to interfere on process B (this would imply A > B) and to be interfered by process B (this would imply A < B), and reciprocally. As a consequence, the amount of mutual interference observed in the simulations when the two processes were of comparable strength was negligible (see Figure 1, Situation 3). Cohen et al. themselves acknowledge that their difficulty to account for the mutual interference reported by MacLeod and Dunbar is a robust property of their model, and they construed this outcome as a shortcoming (p. 353).

Is this shortcoming really devastating for the validity of Cohen et al.'s (1990) model? One may reasonably think that a single empirical demonstration of bidirectional interference (MacLeod & Dunbar, 1988) does not strike a fatal blow to the model. Unfortunately, as shown in the prior section, the other relevant studies do not clearly confirm or disprove this demonstration. A small amount of mutual interference is sometimes reported in quite specific conditions, a form of evidence that moderately challenges the Cohen et al.'s model without providing a definitive case against it. The present study was devised to clear up the current ambiguities.

The present study

As described above, replacing colors with musical notes as the irrelevant dimension in a reading task resulted in the observation of a reliable RSE in musicians (Grégoire et al., 2013). Reading times were slowed down by incongruent note positions, even though note names were not degraded and were printed in a font that certainly surpassed in size the fonts used in most books or newspapers. However, as such, these results provide only weak arguments against the unidirectionality of interference. First, the effects were small in amplitude (within a 8-10 ms range).[12] The first aim of the present study was to replicate and strengthen Grégoire et al.'s earlier evidence for an MSE, by using a somewhat simplified procedure that will be described later. To anticipate, the results clearly fulfilled our

[12] In addition, Zakay and Glickson (1985) used similar conditions as a part of a more general paradigm, and verbal responses showed no reliable evidence for an MSE. However, methodological limitations are likely responsible for this failure, as analyzed in Grégoire et al. (2013).

expectations in this regard. Second and above all, Grégoire et al. provided no evidence for mutual interference, given that their experiments did not include a task in which reading would have been the irrelevant dimension.

In the following experiments, musicians were exposed to the congruent and incongruent note name/note position patterns used in Grégoire et al. (2013), but they received two types of instructions in succession: In addition to being asked to read the word while ignoring the location of the note within the staff, as in Grégoire et al., musicians were asked to name the note while ignoring the note name written inside the note picture. This latter condition was devised to generate a "reverse MSE" (hereafter: RMSE).[13] In addition, to assess the relative strength of the processing pathways, musicians were submitted to a reading-ability test and a note-naming ability test after the experimental sessions. The note-naming test was also exploited to check whether all musicians reached a sufficient level of musical expertise.

Method

Participants. All participants were undergraduate psychology students at the University of Bourgogne, who received course credit for their participation. They were French native speakers and reported normal or corrected vision. Twenty-six participants had received formal musical training and had played a musical instrument for at least five years. Four of them were removed from the analyses on the basis of their performance in the note-naming ability test (see detail below). The remaining 22 participants (17 women, 5 men) had an average of 12.50 (*SD* = 4.51) years of musical training. Twenty-two other participants had never studied or practiced music (18 women, 4 men).

Materials. The experimental material was composed of four types of stimuli. For the congruent and incongruent conditions, stimuli consisted of a treble staff with a note picture, which could appear on each of the 13 possible positions going from C4 to A5. The name of a note was written inside the note picture. For the congruent condition, the note name was congruent with the note position on the staff (Figure 2a), whereas in the incongruent condition, note name and position were incongruent, with the name written inside the note picture being one of the six other possible note names (e.g., when the note was DO, the

[13] Given that the MSE is an instance of RSE as defined above, the RMSE, in which reading is the interfering process, has the status of a Stroop-like effect. This terminological quibble is unfortunate, but seems unavoidable, because it is difficult to revisit a terminology forged through eight decades of tradition during which reading has been the almost exclusive target in Stroop studies.

written name could be LA, SI, RE, MI, FA, or SOL; Figure 2b).

The neutral condition was aimed at assessing whether the overall difference between congruent and incongruent trials was due to interference or facilitation. For the word-reading task, a note name was inserted inside a note picture, and the resulting pattern was displayed on the same 13 spatial locations as for congruent and incongruent trials (from C4 to A5). However, the staff was not represented (Figure 2c). The stimuli described so far were identical to those used in Grégoire et al. (2013). For the note-naming task, a note picture was presented in a treble staff, with *XX* or *XXX* written inside (Figure 2d). Note pictures were also displayed at all 13 note locations.

Grégoire et al. (2013) also used stimuli that were made up of words that were not names of notes (and hence neither congruent nor incongruent with note locations), but instead highly frequent French words that were matched in length with note names. The results suggested that these additional conditions were not necessary in further investigations, and the authors conjectured that using only the experimental conditions involving note names would result in larger congruity effect. Indeed, as a by-product of mixing note names with other words, the proportion of congruent trials was reduced (20%), and it has been shown in the standard color-word version that interference increases with the proportion of congruent trials (Lowe & Mitterer, 1982). In the present experiment, only note names were used.

The stimuli used in the independent tests of reading and note-naming abilities were designed to be closer from those involved in everyday situations. They are represented in Figure 1e and Figure 1f, respectively.

To prevent the iconic memory of the staff to influence the processing of the following note, the stimuli were randomly displayed at one of four possible positions without immediate repetition at the same location. The four positions were defined as the center of (invisible) rectangles resulting from the exhaustive partitioning of the screen into four quadrants of equal size. Stimuli were printed in black over a white background on a computer screen. Note names and Xs appeared in a font that roughly corresponds to the "Arial 14", uppercase font. The treble staff was 7.7 cm wide by 5.1 cm high.

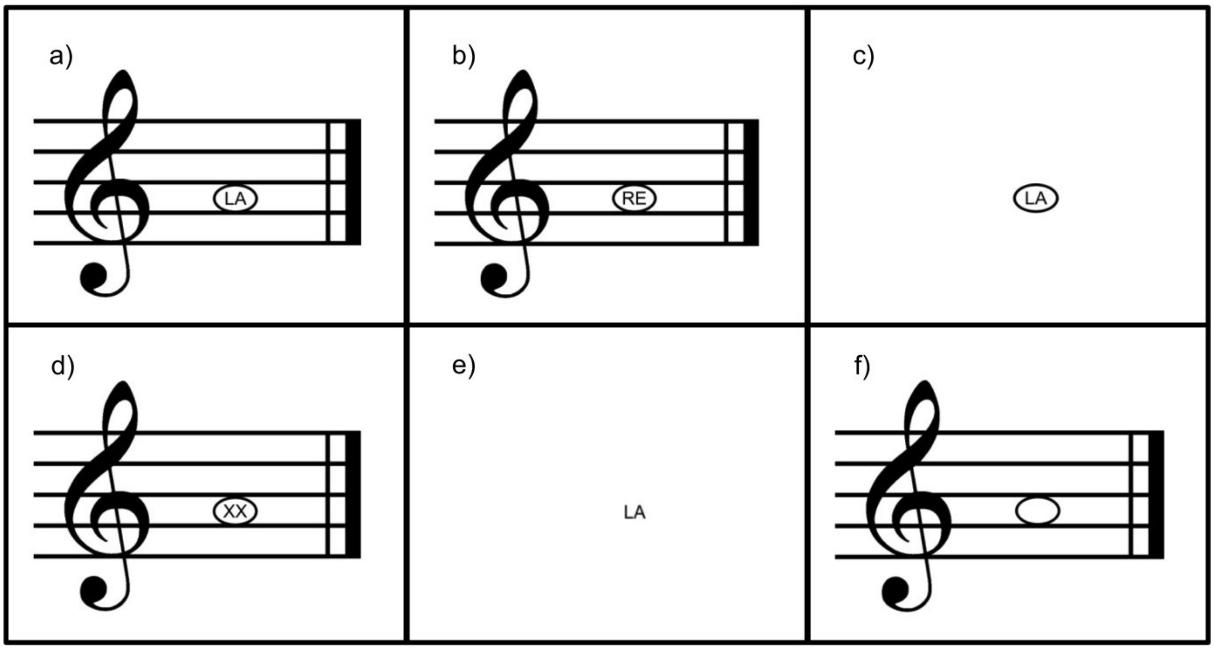

Figure 2. Examples of the different conditions used in the experiment: a) congruent condition; b) incongruent condition; c) neutral condition for word reading; d) neutral condition for note naming; e) reading-ability test; f) note-naming ability test. Note that in the musical French notation (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

Procedure. Musicians had to perform a word-reading task and a note-naming task in succession. The order of presentation of the two tasks was counterbalanced across participants. In the word-reading task, participants had to read aloud the printed word while ignoring the note picture. In the note-naming task, musicians were asked to name the note while ignoring the word written inside. Nonmusicians did not perform the note-naming task.

For each task, there were three mixed conditions: congruent (Figure 2a), incongruent (Figure 2b), and neutral (Figure 2c, d). For each condition, the stimuli appeared six times on each of the 13 locations, resulting in 78 trials per condition, and 234 trials (78 × 3) for each task. On each trial, a fixation cross was displayed for 1 s at the center of the screen before the apparition of the stimulus, which stayed on the screen until participant's response. The interval between the response and the next trial was 1s. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of the same response. They were displayed as six blocks of 39 trials each with a self-paced break between blocks.

For the musicians, the experimental session was immediately followed by two additional tests, which were run in counterbalanced order. One test was a reading-ability test, in which participants had to read note names (Figure 2e). The other test was a note-naming ability test, in which participants had to name notes (Figure 2f). Each test included 78 trials. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of words and notes. They were displayed as two blocks of 39 trials each with a self-paced break between blocks

Whatever the tasks, participants were encouraged to respond as fast and as accurately as possible throughout the session. The response times (RTs) were recorded by a voice key. During the session, the experimenter noted the error responses and the voice-key dysfunctions. After the experiment, the musicians filled out a questionnaire about their musical training.

Results

Ability tests (musicians)

Four musicians were removed from the analyses on the basis of their poor performance in the note-naming ability test: Three of them took longer than 1000 ms on average to name a note, and a last one made seven errors (a majority of the remaining 22 musicians made no error, and their highest error score was four).

Voice-key dysfunctions led to exclude 3.06% of the data. Unsurprisingly, there was virtually no error in the reading-ability test, M = 0.12%, SD = 0.39. The proportion of errors in the note-naming ability test was significantly higher, M = 1.47%, SD = 2.05, $t(21) = 3.10$, $p = .005$, $d = 0.94$. RTs for correct responses beyond three standard deviations of the overall mean per participant (0.73%) were removed. Correct RTs were also significantly shorter for word reading than for note naming, M = 615.37, SD = 62.30 vs. M = 742.28, SD = 94.33, respectively, $t(21) = 8.21$, $p < .001$, $d = 1.79$. Note that there was no evidence for a speed-accuracy trade-off in the note-naming task, as shown by the positive (although nonsignificant) correlation between errors and RTs, $r(20) = .21$, $p = .35$.

Word reading (MSE)

Voice-key dysfunctions led to exclude 4.28% of the data. While nonmusicians made no reading errors, a small proportion of errors was observed for musicians (0.75%). An ANOVA on musicians' reading errors performed with Condition (congruent, incongruent, neutral) as a within-subject variable showed a significant main effect of condition, $F(2, 42) = 4.88$, $p = .012$, $\eta_p^2 = .189$. The rate of errors was significantly higher in the incongruent condition (M = 1.89%, SD = 3.76) than in the congruent condition (M = 0.06%, SD = 0.27), $t(21) = 2.30$, $p = .032$, $d = 0.50$, attesting to an MSE. The rate of errors in the neutral condition (M = 0.30%, SD = 0.68) was both lower than the rate of errors in the incongruent condition, $t(21) = 2.11$, $p = .047$, $d = 0.46$ and higher than the rate of errors in the congruent condition, $t(21) = 2.16$, $p = .042$, $d = 0.47$.

RTs for correct responses beyond three standard deviations of the overall mean per participant (0.59%) were removed. The remaining data are shown in Figure 3. An ANOVA on RTs was carried out with Condition (congruent, incongruent, neutral) as a within-subject variable and Musical Expertise (musicians, nonmusicians) as a between-subject variable. There was no main effect of musical expertise, $F(1, 42) = 0.52$, $p = .474$, a significant effect of condition, $F(2, 84) = 52.94$, $p < .001$, $\eta_p^2 = .558$, and more importantly, a significant Condition × Musical Expertise interaction, $F(2, 84) = 4.54$, $p = .013$, $\eta_p^2 = .098$.

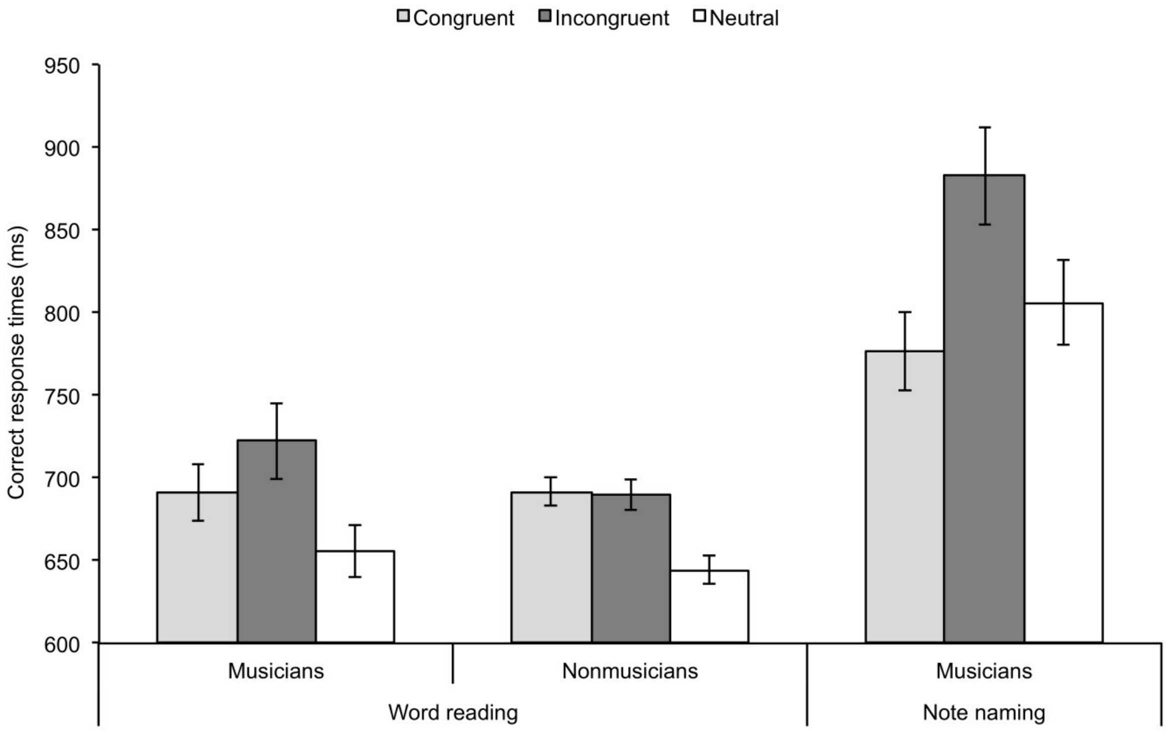

Figure 3. Correct response times as a function of Task, Condition, and Musical Expertise. Error bars indicate standard errors.

For Musicians, RTs were significantly longer in incongruent condition than in congruent condition, $t(21) = 3.21$, $p = .004$, $d = 0.70$. While the MSE reported by Grégoire et al. (2013) was in the 8-10 ms range, the difference reached 31.36 ms in this experiment, presumably due to the removing of control conditions. Two questions arise regarding the generality of the effect: First, is it observed for all participants? And second, for a given participant, is it observed across the entire latency distribution of his/her responses? Regarding the first issue, RTs were numerically longer in incongruent condition than in congruent condition for 20 out of the 22 musician participants. The second issue was addressed by calculating so-called cumulative Vincentized distribution functions (Ratcliff, 1979; Roelofs, 2008). The rank-ordered RT distribution was divided into deciles (10% quantiles) for each participant separately for the congruent and the incongruent conditions, then the data were averaged across participants. The resulting curves are reported in Figure 4, upper panel. The MSE, which is represented on the figure as the horizontal distance between the curves, was stronger when RTs were longer, generating a significant interaction between congruity (congruent, incongruent) and deciles (1-10), $F(9, 189) = 10.95$, $p < .001$, $\eta_p^2 = .343$. Importantly, despite this effect, individual *t*-tests showed that the effect of conditions was significant even for the shortest RTs (first decile: $p = .045$, second decile: $p = .031$, other deciles: all $p < .01$).

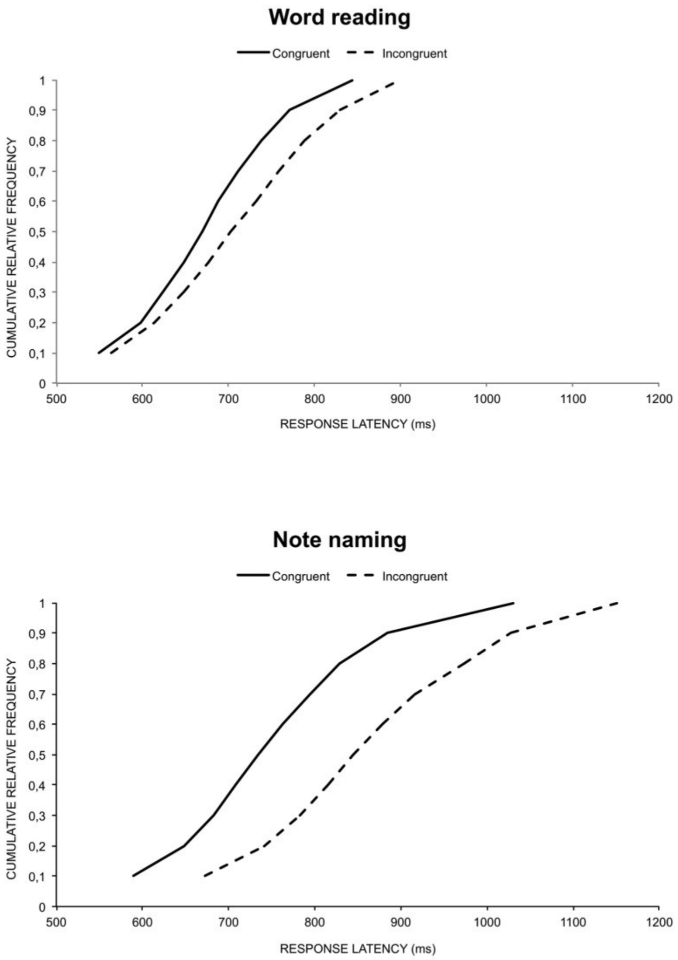

Figure 4. Vincentized cumulative distribution curves for word reading and note naming in the congruent and incongruent conditions.

RTs did not differ between incongruent and congruent conditions for the nonmusicians, $t(21) = 1.06$, $p = .301$. Comparing the effect between musicians and nonmusicians revealed a significant difference, $t(42) = 3.36$, $p = .002$, $d = 1.04$, attesting that the MSE was specific to musicians.

Grégoire et al. (2013) showed that the neutral word condition is not appropriate to provide a direct measure of facilitation and interference, because perceptual complexity is lower in this condition than in the congruent and incongruent conditions. Nevertheless, the effect of perceptual complexity can be assessed in nonmusicians, as a difference between the congruent and incongruent conditions (which gave nearly identical RTs) on the one hand, and the neutral condition on the other hand. The resulting value was 46.26 ms. If one supposes that musicians and nonmusicians are equally sensitive to perceptual complexity, this value can be added to the value observed for musicians in the neutral condition to obtain a neutral value that incorporates the effect of perceptual complexity. This corrected neutral value would thus be 701.95 ms (655.69 + 46.26). Given that the RTs in congruent and incongruent conditions for musicians were 691.16 ms and 722.52 ms, respectively, the overall difference (31.36 ms) may be decomposed into a facilitation effect (701.95 - 691.16 = 10.79 ms) and a larger interference effect (722.52 - 701.95 = 20.57 ms). Note that this analysis rests on the postulate that musicians and nonmusicians are equally sensitive to perceptual complexity, which is somewhat unrealistic. Most probably, musicians are less sensitive than nonmusicians to the complexity of a picture they are highly familiar with. This implies that the corrected neutral value should be lower, and hence that the part of facilitation in the analysis above is certainly overestimated. This descriptive analysis confirms Grégoire et al.'s observation that the MSE is mainly due to interference in the incongruent condition.

Note naming (RMSE)

Voice-key dysfunctions led to exclude 5.63% of the data. An ANOVA on naming errors performed with Condition (congruent, incongruent, neutral) as a within-subject variable (recall that only musicians performed this task) showed a significant main effect of Condition, $F(2, 42) = 32.13$, $p < .001$, $\eta_p^2 = .605$. Musicians made significantly more errors in the incongruent condition ($M = 7.10\%$, $SD = 4.56$) than in the congruent condition ($M = 0.86\%$, $SD = 1.79$), $t(21) = 6.03$, $p < .001$, $d = 1.32$, and than in the neutral condition ($M = 1.64\%$, $SD = 2.92$), $t(21) = 5.80$, $p < .001$, $d = 1.27$. Although there were more errors in the neutral condition than in the congruent condition, the difference failed to reach significance, $t(25) = 1.75$, $p = .095$, $d = 0.38$.

Correct RTs beyond three standard deviations of the overall mean per participant (1.21%) were removed. The remaining data are shown in Figure 3, right panel. As for the errors, an ANOVA carried out on the correct RTs with Condition (congruent, incongruent, neutral) as a within-subject variable revealed a significant main effect of Condition, $F(2, 42) = 64.49$, $p < .001$, $\eta_p^2 = .754$. RTs in the incongruent condition were significantly longer than in the congruent condition, $t(21) = 9.20$, $p < .001$, $d = 2.01$, hence attesting to the presence of an RMSE. This effect was present for all musician participants, and cumulative Vincentized distribution functions calculated as above (Figure 4, lower panel) showed that the effect was observed throughout the entire RT range ($p < .001$ for all deciles). As for the MSE, the RMSE was stronger when RTs were longer, generating a significant Congruity × Decile interaction, $F(9, 189) = 9.38$, $p < .001$, $\eta_p^2 = .309$. In comparison with the neutral condition, we observed an interference in the incongruent condition, $t(21) = 7.06$, $p < .001$, $d = 1.54$, and a facilitation in the congruent condition, $t(21) = 5.56$, $p < .001$, $d = 1.21$.

Are the effects constant throughout the test sessions?

In most Stroop paradigms including the present experiment, there is the same number of congruent and incongruent items in order to comply with standard methodological requirements. However, by construction, the number of different incongruent items exceeds the number of different congruent items. As a consequence, congruent items are more often repeated than incongruent items, hence making it possible to learn from the sequence of items (see Melara & Algom, 2003). In the present experiment, the large number of different congruent items (i.e., 13) makes learning from the test somewhat unlikely. In addition, if some learning occurred during the experiment, this should be true for both musicians and nonmusicians. Therefore, the performance of nonmusicians should have departed from chance, which is not the case.

However, it remains possible that congruity effects increased throughout sessions for musicians, or still decreased, for instance due to the gradual emergence of strategic factors. To address this issue, for each task, we divided the whole session ($N = 234$) into three blocks of equal length ($N = 78$), and the effect of congruity was computed for each block as the mean differences between scores for incongruent and congruent trials. There was no reliable change across sessions, whatever the task (note naming or word reading) and the dependent variable (errors or RTs) were. For the errors, an ANOVA with Task (note naming, word reading) and Block (1, 2, 3) as within-subject factors returned a main effect of Task (see the next section for further analyses), but no main effect of Block, $F(2, 42) = 0.45$, $p = .642$, and no Task × Block interaction, $F(2, 42) = 0.16$, $p = .851$. For RTs, the same analysis also gave a main effect of Task (see the next section for further analyses), but no main effect of Block, $F(2, 42) = 0.31$, $p = .738$, and no Task × Block interaction, $F(2, 42) = 1.44$, $p = .248$. Overall, these results indicate that the effects observed in musicians were a stable reflect of the expertise musicians have gained from their musical training in everyday life.

Comparing MSE and RMSE

An ANOVA on musicians' errors with Condition (congruent, incongruent, neutral) and Task (note naming, word reading) as within-subject variables revealed significant main effects of Condition, $F(2, 42) = 27.24$, $p < .001$, $\eta_p^2 = .565$, and Task, $F(1, 21) = 18.57$, $p < .001$, $\eta_p^2 = .469$, and a significant Condition × Task interaction, $F(2, 42) = 13.86$, $p < .001$, $\eta_p^2 = .398$. The congruity effect, as assessed by the difference between incongruent and congruent condition, was also stronger for the note-naming task ($M = 6.24\%$, $SD = 4.85$) than for the word-reading task ($M = 1.83\%$, $SD = 3.74$), $t(21) = 3.86$, $p = .001$, $d = 0.84$.

The analyses ran on musicians' RTs returned the same results, with significant effects of Condition, $F(2, 42) = 68.39$, $p < .001$, $\eta_p^2 = .765$, and Task, $F(1, 21) = 26.93$, $p < .001$, $\eta_p^2 = .562$, and a significant Condition × Task interaction, $F(2, 42) = 13.98$, $p < .001$, $\eta_p^2 = .40$. The congruity effect was much stronger for the note-naming task than for the word-reading task (106.30 ms vs. 31.36 ms respectively), $t(21) = 4.47$, $p < .001$, $d = 0.98$. To sum up, the RMSE exceeded the MSE with both accuracy and RT measures.

Individual Differences

In the Cohen et al.'s (1990) model, the pattern of interference depends on a single parameter: the relative strength of the two competing pathways. If one postulates that RTs in the post-experimental ability tests provide a measure of strength (see below for a discussion), the relative strength of the word-reading and note-naming pathways for a given musician can be given by the difference of RTs in the two tasks (as mentioned above, error rates were negligible in the ability tests). Although naming notes took longer than reading for all musicians, the difference largely differed across musicians (range: from 17.61 ms to 230 ms). A participant P1 whose difference between note naming and word reading is larger than for a participant P2 can be taken as having a stronger imbalance between musical and reading abilities than participant P2. Is this difference actually predictive of the pattern of interference?

The response is clearly positive. There was a negative correlation between the relative strength of the two pathways and the amount of interference in the word-reading task, $r(20) = -.524$, $p = .012$, and there was a positive correlation between the relative strength of the two pathways and the amount of interference in the note-naming task, $r(20) = .428$, $p = .047$. To provide a more complete picture of the resulting pattern, participants were divided into three (roughly equal) groups along the relative strength dimension, with 7, 8, and 7 participants in each group. At one extreme of the resulting classification, the mean difference in speed between the two pathways was moderate (range: 17.61 - 76.58 ms). Participants exhibiting this pattern will be called *more balanced*. At the other extreme, the difference in strength between the two pathways is maximal (i.e., note naming is much longer than word reading, range: 181.06 - 230.00 ms). Participants exhibiting this pattern will be called *less balanced*. The remaining participants (*intermediate*) are in-between.

Figure 5 shows the pattern of interference for each group. Overall, an ANOVA on the amount of interference with Group as a between-subject variable (more balanced, intermediate, less balanced) and Task (word reading, note naming) as a within-subject variable revealed no main effect of Group, $F(2, 19) = 0.28$, $p = .756$, a significant effect of Task, $F(1, 19) = 24.60$, $p < .001$, $\eta_p^2 = .564$, and, crucially, a significant Group × Task interaction, $F(2, 19) = 3.72$, $p = .043$, $\eta_p^2 = .281$. For the more-balanced group, the RMSE and the MSE were both different from zero, $t(6) = 2.75$, $p = .033$, $d = 1.24$, and $t(6) = 3.03$, $p = .023$, $d = 1.12$, respectively, and their amplitude did not significantly differ, $t(6) = 0.45$, $p = .672$. By contrast, for the less-balanced group, the RMSE substantially exceeded the MSE, $t(6) = 6.68$, $p < .001$, $d = 2.73$, which was no longer significant, $t(6) = 1.76$, $p = .128$. For the intermediate group, the RMSE significantly exceeded the MSE, $t(7) = 5.18$, $p = .001$, $d = 1.96$, but the MSE was still reliable, $t(7) = 2.45$, $p = .044$, $d = 0.93$.

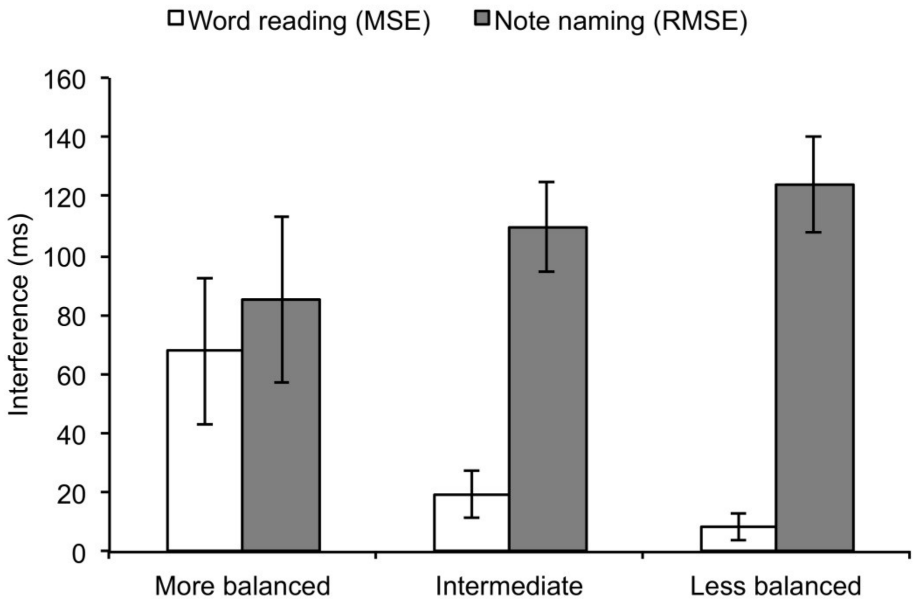

Figure 5. Amount of interference (RTs in incongruent condition minus RTs in congruent condition) in the word-reading and note-naming tasks for three groups of musicians, sorted according to their performance in independent tests of reading and note-naming ability. More-balanced group: note naming was only slightly longer than word reading; Less-balanced group: note naming was largely slower than word reading; Intermediate group: in-between. Error bars indicate standard errors.

The same pattern was observed when the amount of interference was assessed through the rate of errors. As for RT measures, there was a negative correlation between the relative strength and the amount of interference in the word-reading task and a positive correlation between the relative strength and the amount of interference in the note-naming task. However, presumably due to the small number of errors, the correlations did not reach significance, $r(20) = -.321$, $p = .146$, and $r(20) = .347$, $p = .114$, respectively.

Discussion

Grégoire et al. (2013) reported a Musical Stroop Effect (MSE), whereby musicians showed slightly longer times for reading the name of a note printed in incongruent position on a staff than for reading the same note name in congruent position. The first contribution of the reported experiment was to confirm and strengthen this evidence, by showing a much larger interference of reading by note naming in a simplified paradigm. The second and main contribution of the experiment was to demonstrate that the same participants also exhibited interference of note naming by reading. Naming a note took longer when the note name written inside the note was incongruent with the note location on the staff than when it was congruent, giving evidence of a reverse MSE (RMSE). Incongruent patterns also generated significant increases of errors for both the MSE and the RMSE.

As pointed out in the introduction, reports of a reverse effect are not infrequent. However, the reverse effect is obtained in especially designed conditions, including the size and the orientation of the words (e.g., Palef & Olson, 1975), the level of training on a specific dimension (MacLeod & Dunbar, 1988), and the selection of participants (Akiva-Kabiri & Henik, 2012). The crucial point is that, in most cases, these conditions turned out to be inappropriate for the regular effect. Admittedly, a few studies (Melara & Mounts, 1993; Palef & Olson, 1975) have reported bidirectional effects without varying the conditions, but these effects were described as small and fleeting, and could be due, for instance, to the mixture of unidirectional effects going in opposite directions for different participants (Cohen et al., 1990). As a consequence, the bulk of the evidence from the earlier literature is consistent with the conclusion that interference operates in a unidirectional way, which may alternate as a function of conditions or participants.

Our results are clearly at odd with this conclusion. The MSE and the RMSE were obtained in the very same conditions, which were chosen to match as best as possible the usual conditions of reading and note naming in real word settings. Moreover, 20 out of the 22 musicians showed the two effects on RTs (at a numerical level), excluding the possibility that bidirectionnal effects emerged from an artifact due to group statistics. Likewise, analyses of Vincent curves gave evidence that both the MSE and the RMSE were significant across the full range of latencies of the responses, even though the effects were stronger for the slower responses.[14] Finally, the idea that getting bidirectional effects could be dependent on the fact that all our participants were precisely at some optimal level of automatization in the two tasks (e.g., as in Macleod & Dunbar, 1988) is quite implausible: The selection criterion for our musicians was sufficiently loose (a minimum of 5 years of academic teaching) to allow the recruitment of participants largely differing on their level of musical proficiency. To conclude, the present study reports the strongest evidence to date for mutual interference between two competing processes.

The remainder of this paper first examines the implications of these results for the Cohen et al.'s (1990) model, the main principles of which were described in the introduction. Then we address the question of why the results collected with note naming in this paper

[14] Some proportionality between the size of a difference and the raw values on which this difference is computed is a common observation. In addition, the effect may be strengthened by the fact that slow responses may be less automatized, and hence, following Cohen et al.'s model, more receptive to interference, than fast responses.

differ so strikingly from the results observed with color naming or picture naming in the earlier literature.

The Cohen et al.'s (1990) model

We pointed out above that Cohen et al. acknowledged that their difficulty to account for the mutual interference reported by MacLeod and Dunbar (1988) was a shortcoming of their model. However, insofar as mutual interference was only observed in a specific condition (at an intermediary level of training) by MacLeod and Dunbar, and looked as a nearly anecdotal phenomenon in the Stroop literature, the challenge seemed of limited importance. On the face on it, our results change the landscape and drastically increase the charges against the model.

However, we would like to argue that the problem raised by the observation of mutual interference is linked to the specific connectionist implementation and that, paradoxically, our data provide a striking support for the underlying psychological principles of the model. As noted in the introduction, Cohen et al. postulated that the strength of a pathway both increases its ability to produce interference and reduces its susceptibility to interference. In the connectionist structure, this property generates mutual exclusivity of interference when two processes compete, because the stronger process both interferes on the weaker process and is immune to the interference the weaker process could produce. An effect or its reverse is obtained as a function of which of the two processes is the stronger. This implication does not seem to be a principled constraint, however. Rather than reasoning as if a change in the relative strength of the processes resulted in an all-or-none inversion in the direction of interference, the consequence could be conceived of as a gradual modification as well. The Cohen et al.'s claim, according to which it is "the relative strength, compared with a competing pathway [that is] important in determining whether a process will produce or be subject to interference in a Stroop-like task" (p. 348) could thus be rephrased into a softer version: The relative strength, compared with a competing pathway would be important in determining the relative propensity of the stronger process to interfere on, and be interfered by another, weaker process; the final pattern taking the form of a trade-off between the amount of interference affecting each pathway (see Figure 1, for a schematic representation of the suggested modification to the Cohen et al.'s model).

This softened version accounts for the whole pattern of results. The strength of word reading, as assessed by the speed of processing in the ability test, was considerably larger than the strength of note naming in musician participants. In keeping with the reasoning above, the stronger process, namely reading, elicited much larger interference on note naming than the reverse. But the most striking evidence came from the analyses of individual differences. The extent to which the strength of word reading exceeded the strength of note naming in musicians correlated negatively with the interference in word reading (MSE), and positively with the interference in note naming (RMSE). As shown in Figure 5, the subgroup of musicians who exhibited the best balanced performances in the naming test and the reading test also exhibited the best balanced amount of mutual interference, with the MSE being statistically indistinguishable from the RMSE. At the other end, the subgroup of musicians who exhibited the strongest imbalance in the naming test and the reading test also exhibited the strongest asymmetry of the effects, with the MSE being ultimately not statistically different from chance (note that these analyses were performed on very small samples of participants, which implies that statistical nonsignificance has only an indicative value). To sum up, the relative strength of the two competing pathways determined the pattern of interference, exactly as (a somewhat softened version of) Cohen et al.'s model would predict.

The analyses above take for granted that Cohen et al.'s concept of strength may be estimated by the speed of processing, and it could be argued that this postulate is questionable. Cohen et al. mentioned repeatedly that speed must be considered jointly with the amount of practice and the pattern of interference to assess strength. This argument calls for several comments. (1) Assuming that the duration of practice in a natural environment could be estimated with some confidence, directly comparing these durations between reading and note naming would be nearly meaningless. Indeed, the level of automaticity resulting from a same amount of practice obviously depends on the processes involved in the practiced task. For instance, Logan and Klapp (1991) have shown that genuine properties of automaticity can emerge after less than 15 min of training on a simple and consistent alphabet-arithmetic task, an amount of practice that is not in the order of magnitude required to master complex activities such as reading. (2) Exploiting the pattern of interference to assess the strength of the pathways would be obviously circular in our case, given that we are trying to exploit the notion of strength to account for this pattern. (3) Although Cohen et al. took care of distinguishing *strength* and *speed*, they also noted, for instance, that the smaller effects of words on picture naming than on color naming "does not seem to be due to a difference in strength between picture naming and color naming, because both have comparable reaction times in the control condition" (their footnote 18). This suggests that approximating the strength of a pathway through speed measures is not a so strong circumvention of the Cohen et al.'s framework. (4) It is worth stressing that exploiting speed does not amount to endorse the so-called horse-race model of the Stroop effect. The claim that the direction of interference depends on which of the two competing processes is completed first has been clearly rejected, among others by Glaser and Glaser (1982) and Dunbar and Macleod (1984) studies, but this does not entail to discredit speed as a useful correlate of automaticity, at least under specific conditions.

Finally, our assessment of Cohen et al.'s concept of *strength* through speed measures finds support in a last, more complex but also more fundamental argument. The main problem of using speediness to measure strength is the same as for using the amount of practice: All depends on the processes involved in the task, hence making between-task comparisons hazardous. In this regard, the claim above that reading is stronger than note naming because it is faster is admittedly questionable. But it is worth stressing that the part of evidence stemming from the analyses of individual differences does *not* face the same objection. Only within-task comparisons are exploited in these analyses. The only prerequisite is that *for a given task*, speed measures allow comparing the level of strength of the underlying processes between different participants. For instance, a participant P1 who reads notes faster than a participant P2 is taken as having a stronger pathway for note naming than P2, and likewise for reading. Now, a participant P1 whose the difference between note naming and word reading is larger than for a participant P2 can be taken as having a stronger asymmetry between musical and reading abilities than participant P2. This line of reasoning does not require that the absolute value of the difference between the speed of word reading and note naming makes sense.

To conclude, our findings have contrasted implications for the Cohen et al.'s model. On the one hand, the observation of mutual interference for nearly all musicians despite their unequal ability to name notes runs clearly against the connectionist implementation of the model, and presumably implies deep structural changes. On the other hand, however, our data provide a powerful support to the model intuition that the strength of a pathway both increases its ability to produce interference and reduces its susceptibility to interference, and that the final pattern of interference between two competing processes depends on their relative strength.

What is special about note naming?

So far, our analysis has focused on the ability of Cohen et al.'s (1990) model to account for the possibility of mutual interference. However, our results also raise a related, but different question: Why did these results depart from the data collected in earlier studies? In particular, why did note naming turn out to interfere with reading, whereas neither color naming nor picture naming are able to do so without a strong perceptual degradation of the printed materials?

The Cohen et al.'s model does not offer a clear response to this question. The model suggests that relative strength could be responsible, with note naming in musicians having greater strength than color or picture naming. However, to avoid circularity, this interpretation should be validated through independent measures of strength, and this issue remains unsettled. Comparing the level of practice in natural settings between note naming on the one hand, and color naming or picture naming on the other hand, makes little sense, for reasons discussed above. Moreover, measuring strength through the speed of responses does not support the hypothesis: RTs in the test of musical reading ability were not shorter, and in fact numerically longer (mean RT = 742 ms in our experiment) than those commonly reported for color naming and picture naming, both of which are comparable (Cohen et al. estimate the mean RT to name color or picture to approximately 650 ms, see their footnote 18). Note that these chronometric data also lead to disregard models relying directly on the relative speed of processing (e.g., Logan, 1980), as a possible explanation for the specificity of note naming with regard to color or picture naming.

Melara and Algom (2003; see also Melara & Mounts, 1993) have questioned the relevance of a direct comparison between the strength (Cohen et al., 1990) or the speed (Logan, 1980) of the competing processes, as classically evaluated by the strength or speed of processing of individual stimulus attributes. They suggest rather that the relevant variable is the perceptual discriminability of the stimuli as assessed within each dimension. For each dimension, the level of discriminability is measured in the experimental context, by the speed or the accuracy of identification of two values alternating randomly along this dimension from trial to trial. The direction of the interference would depend on the relative discriminability of the values for the two competing dimensions, with the more discriminable dimension disrupting classification of the less discriminable dimension. Although these ideas have received some experimental support, their adequacy to account for our data is questionable. For our concern, at least two conditions should be fulfilled. The first is that the discriminability of musical notes (for musicians) should be much better than the discriminability of the colors or pictures used in standard Stroop tasks. Although further empirical studies would be needed at this point, we see no a priori reason that would lend support to this conjecture. Second and more importantly, one needs to posit that once the condition of matched discriminability is reached (or at least approximated), a sizeable mutual interference follows. The Melara and Algom’s (2003) model does not make this prediction, at least explicitly. At an empirical level, when discriminabilities were matched, Melara and Mounts (1993) observed "small interactive effect being reduced or eliminated through practice" (their abstract). Overall, irrespective of the interest of the Melara and collaborators' framework, it looks unlikely that it is appropriate to explain our results.

At first glance, WEAVER++, the model proposed by Roelofs (2003, see also Roelofs, 1997, 2005), is more promising. Indeed, WEAVER++ successfully accounts for the mutual interference observed in MacLeod and Dunbar (1998) at an intermediate stage of practice. The predictions of WEAVER++ are grounded on the prior knowledge of the task architectures, gained by independent evidence coming from behavioral and neuropsychological investigations. The asymmetry between reading and color naming is explained within the theory of Levelt, Roelofs, and Meyer (1999), according to which a process of conceptual identification would precede articulation in color naming, whereas the order of these two steps of processing would be reversed for reading aloud. As a consequence, color would require an extra processing step before reaching articulation compared to word.

The Roelofs' model predicts mutual interference whenever two tasks involve the same architecture. The explanation for mutual interference in MacLeod and Dunbar (1998) relies on the fact that none of the tasks used in this study—color naming and shape naming—involved the reading architecture. Given that the reading architecture is necessarily involved as one of the two dimensions in our procedure, the reason why WEAVER++ achieves to account for mutual interference in MacLeod and Dunbar cannot be transposed to our experiment. The only solution to account for our results in the Roelofs' model would consist in attributing to note naming the same architecture as for reading, or at least an architecture that would be closer to reading than to color naming or picture naming. Our lack of background knowledge on the architecture of the note-naming pathway makes further speculations hazardous, but we submit nevertheless that this condition seems to be hard to meet. Indeed, on the face of it, note naming is nothing else but an instance of picture naming (note we are dealing here only with note naming such as involved in our experiment, and not with the full array of processes involved in music reading). A possibility would be that with very extensive training, the extra processing step involved in note naming turns out to be by-passed. Further studies on the psychological processes involved in note naming in expert musicians are needed before firmer conclusions may be drawn.

Another popular model of the Stroop effect is the translational account (e.g., Durgin, 2003; Virzi & Egeth, 1985). This model is aimed at accounting for the asymmetry commonly reported in the literature between reading and any other competing dimension. The underpinning idea is that interference arises only if responding to the target implies a translation from one code to another code. For instance, naming a color implies for the color code to be converted into a verbal code, and this translation makes the vocal response vulnerable to interference. By contrast, the written word is already in the verbal code serving for the response and hence, reading aloud is immune to any interference.

As such, the translational model is obviously unable to account for the part of our results evidencing interference on reading, and in this regard, the reported evidence for an MSE sounds as a strong case against the model (for other contradictory data, see Blais & Besner, 2007). Could the translational model be completed or modified to account for mutual interference nevertheless? The only solution we see as compatible with the model would be to add the postulate that two competing processes involving no translation could exhibit mutual interference. To work in the present situation, this would need to consider that note naming implies no translation, hence differing in this regard from both color naming and picture naming. Unfortunately, naming a note seemingly implies a translation from a pictorial code to a verbal code, in the very same way as naming a picture.

To conclude, our examination of different concepts involved in the current Stroop models to account for the direction of interference—strength, speed of processing, perceptual discriminability, task architecture, or still code translation—leaves us rather somewhat pessimistic on their propensity to explain the particularity of note naming with regard to the more common dimensions competing with reading in Stroop paradigms.

Reconsidering an old idea

We suggest that the discrepancy between our and earlier results could be explained in reference to the early observation of Peterson, Lanier, and Walker (1925; cited in MacLeod, 1991; Peterson was the supervisor of John Ridley Stroop's thesis). Peterson et al. wrote: "To the written words "red", "blue", "green" etc., the subjects have as a rule given in the past but the one response of pronouncing (vocally or subvocally) the names of these colors; whereas on seeing the colors red, blue, green, etc. they have responded in many different ways, as grasping and eating, handling, perceiving and admiring, etc. In the case of the words, then, but one specific response-habit has become associated with each word, while in the case of the colors themselves a variety of response tendencies has been developed." (p. 281). Our claim is that note naming is much closer to reading than to color or picture naming on that dimension. As a rule, a musician gives a response of pronouncing vocally (e.g., at the beginning practice) or subvocally (e.g., when playing an instrument later in practice) the name of the note when exposed to a note on a staff.

To our knowledge, this idea has never been articulated as a full-blown model in the Stroop literature, although similar suggestions have been done on occasion (e.g., Cohen et al., 1990, their footnote 7). However, there is at least another research domain in which a related idea has been heavily developed. We allude here to the research on associative learning, to which Peterson et al. (1925) refer. Since Rescorla (1968), the associative learning tradition, extended by recent studies on statistical learning (see Perruchet & Poulin-Charronnat, 2012, for linking the two domains together), emphasizes the pervasive role of the *contingency* between events in the strengthening of associative relations. The measures of contingency differ from the simple frequency of co-occurrences: The level of contingency indicates how an event is *predictive* of another event. Note that Shiffrin and Schneider (1977, see also Schneider & Shiffrin 1977) have provided impressive evidence for the role of contingency in the formation of automatisms, although they did not used the term of contingency. They refer to the *consistent mapping* of stimulus-response pairs, by contrast to varied mapping, to describe a condition that could also be described as a perfect contingency between events.

To reframe Peterson et al.'s (1925) intuition in contemporary terminology, one could say that, in real-world life, there is a low level of contingency between seeing a color and producing the name of the color. The main reason is *not* that producing the name of a color in response to a colored patch is infrequent: After all, one could argue that children are doing that intensively when learning the colors. The main reason is that colors are continually present in our visual field without their names being evoked, overtly or covertly. The very same comment could be made about picture naming: We are continually faced with objects that, in an overwhelming proportion of situations, we are not naming. By contrast, the contingency between seeing a written word and producing, vocally or subvocally the word is certainly considerably stronger. What about note naming in musicians on this dimension?

Although highly speculative, the idea may be defended that the contingency between perceiving a note on a staff and generating the name of this note is at least equal, if not stronger than the stimulus-response contingency in word reading. Indeed, even for musicians, a note is virtually never displayed in the environment out of the context in which it ought to be named.

If one endorses this framework, the conclusions stemming from the standard paradigms—the color-word and picture-word tasks—would be limited to the case where the process competing with word reading has evolved in conditions that were not conducive to automatization. Using note naming in musicians suggests a new conjecture: When reading competes with another process that is highly automatized due to the strong stimulus-response contingency, getting mutual interference could be the rule.

References

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical stroop effect in absolute pitch possessors. *Experimental Psychology, 59*(5), 272-278.

Blais, C., & Besner, D. (2006). Reverse stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1345-1353.

Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review, 14*(3), 466-469.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332-361.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 470-484.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 622-639.

Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics, 65*(3), 367-378.

Ellis, N. R., Woodley-Zanthos, P., Dulaney, C. L., & Palmer, R. L. (1989). Automatic-effortful processing and cognitive inertia in persons with mental retardation. *American Journal on Mental Retardation, 93*(4), 412-423.

Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in "ignoring" irrelevant information. *Memory & Cognition, 7*(2), 86-94.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 640-654.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*(6), 875-894.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The Musical Stroop Effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*(4), 269-278.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014). Is the musical Stroop effect able to keep its promises ? A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology, 61*(1), 80–83.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-38.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*(4), 523-553.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is Extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 179-195.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided attention in a Stroop task. *Canadian Journal of Psychology, 36*(4), 684-700.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*(2), 201-211.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 126-135.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383-391.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422-471.

Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21*(5), 627-645.

Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology, 29*(3), 201-209.

Perruchet, P., & Poulin-Charronnat, B. (2012). Word segmentation: Trading the (new, but poor) concept of statistical computation for the (old, but richer) associative approach. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 119-143). Berlin: De Gruyter Mouton.

Peterson, J., Lanier, L. H., & Walker, H. M. (1925). Comparisons of white and negro children in certain ingenuity and speed tests. *Journal of Comparative Psychology, 5*, 271-283.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology, 54*(3), 251-282.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446-461.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66*, 1-5.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249-284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*(1), 88-125.

Roelofs, A. (2005). The visual-auditory color-word stroop asymmetry and its time course. *Memory & Cognition, 33*(8), 1325-1336.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(2), 353-368.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing:II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127-190.

Stewart, L., Walsh, & Frith, U. (2004). Reading music modifies spatial mappings in pianists. *Perception & Psychophysics, 66*(2), 183-195.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Verdonschot, R. G., La Heij, W., & Schiller, N. O. (2010). Semantic context effects when naming Japanese kanji, but not Chinese hànzì. *Cognition, 115*, 512-518.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13*(4), 304-319.

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology, 39*(3), 414-423.

Etude 3a

How does Stroop interference change with practice?
A reappraisal from the musical Stroop paradigm

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
CNRS-UMR 5022, University of Bourgogne, Dijon, France

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (submitted). How does Stroop interference vary with practice? A reappraisal from the musical Stroop paradigm.

Author Note

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat, Université de Bourgogne, LEAD/CNRS, UMR5022, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France

This research was supported by the grant ANR-09-JCJC-0129 and the grant Conseil Régional de Bourgogne N° 2012-9201AAO05S01353.

Correspondence concerning this article should be addressed to Laurent Grégoire, LEAD-CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 2100 Dijon, France. E-mail: laurent.gregoire@u-bourgogne.fr

Résumé

La grande majorité des études qui ont investigué l'évolution de l'effet Stroop en fonction du niveau de pratique de la lecture ont reporté une relation en U inversé, selon laquelle l'effet Stroop apparaît tôt pendant l'acquisition de la lecture, atteint un pic vers l'âge de sept ans, et diminue progressivement jusqu'à l'âge adulte. La composante négative de la courbe suggère que les lecteurs expérimentés seraient davantage capables de contrôler leur habileté que les lecteurs moins expérimentés. Néanmoins, dans ces études, le niveau de pratique de la lecture est largement confondu avec l'âge des participants, pour des raisons éthiques et pratiques évidentes. Il est donc possible que la réduction de l'interférence Stroop observée soit due à une capacité croissante des plus vieux enfants à inhiber l'information non-pertinente. Dans la présente étude, la source de l'interférence n'est plus la lecture mais la dénomination de notes chez les musiciens. L'avantage majeur que représente la pratique musicale, par rapport à la lecture, est qu'elle peut être découplée de l'âge. Dans deux expériences exploitant le paradigme de Stroop musical, nous observons une apparition précoce de l'effet Stroop musical (MSE), mais nous ne répliquons pas la relation en U inversé obtenue avec les tâches de Stroop couleur-mot et image-mot. L'expérience 2 révèle une relation linéaire et positive entre l'amplitude du MSE et le niveau de solfège à travers cinq années de pratique musicale chez des enfants âgés de 9-10 ans. Ce résultat suggère que le caractère obligatoire du traitement automatique évolue en parallèle avec les autres propriétés des automatismes, comme la vitesse et l'efficacité.

Abstract

All the earlier studies investigating the evolution of Stroop effect with the amount of reading practice have reported data consistent with an inverted U-shaped curve, whereby Stroop effect appears early during reading acquisition, reaches a peak after 2 or 3 years of practice, then continuously decreases until adulthood. The downward component of the curve suggests that skilled performers would be able to control their performance better than less-skilled performers. However, in these studies, the level of reading practice entirely coincides with age due to obvious practical and ethical constraints, and it is possible that the observed reduction in Stroop interference is due to a growing ability of older children to inhibit nonrelevant information. In the present study, word reading, as source of interference, was traded against note naming in musicians. The major advantage is that musical training can be easily decoupled from age. In two experiments exploiting the musical Stroop paradigm (Grégoire, Perruchet, & Poulin-Charronnat, 2013), we observed an early appearance of the interference effect, as reported for the color-word and picture-word Stroop tasks, but we did not replicate the inverted U-shaped curve. Experiment 2 revealed a linear and positive relation between the amplitude of the musical Stroop effect and the amount of musical practice across five years of musical training. These results suggest that Stroop interference and other measures of automaticity, such as speed and low attentional load, could evolve in parallel with increasing practice.

Introduction

Since the original study of Stroop (1935), in which an incongruent color word was shown to slow down the naming of the ink color in which it was printed (but not vice versa), skilled reading is considered as irrepressible (see MacLeod, 1991; MacLeod & MacDonald, 2000, for reviews). This irrepressibility is obviously dependent on learning literacy, but, surprisingly, only a small proportion of the wide body of literature dealing with the Stroop effect raises the question of how Stroop interference evolves with the amount of reading practice.

An inverted U-shaped function

Comalli, Wapner, and Werner (1962) were the first to investigate this issue. Participants were divided into 11 age groups (7, 8, 9, 10, 11, 12, 13, 17-19, 25-34, 35-44, and a group of older participants which is irrelevant for our concern). Stroop interference was measured as the difference between the time needed to name the color of the ink of incongruent color words (e.g., the word BLUE printed in *green* ink) and the time required to name the color of patches. The greatest Stroop interference was found for seven-year-old children, then the amount of interference decreased up to 17-19 years, and remained constant up to 44 years.

Using the same methodology as Comalli et al. (1962), Wapner, Werner, and McFarland (1963) tested four groups of children (6-7, 9-10, 12-13, and 16-17 years old). In departure to Comalli et al., Stroop interference (as assessed from their Table 10) increased from 6 years old (783 s) to 9 years old (1002 s). However, Rand et al. replicated the decrease in Stroop interference in older children, with a Stroop effect decreasing in amplitude from 9 years old to 12 years old (746 s) and 16 years old (435 s). More recently, Peru, Faccioli, and Tassinari (2006) observed a similar pattern of results in a normative developmental study in Italian, in which children from 3 to 10 years old were tested. Stroop interference increased from 3 to 7 years old (grade 2)[15] and then continuously decreased until 10 years old (grade 5). Schiller (1966; children from grades 1 to 8), Dash and Dash (1982; children from grades 1 to 5) and Schadler and Thissen (1981, nonreaders, beginning readers, children from grades 1 to

[15] Some studies vary chronological age, while others vary school level. For the sake of comparison, 6-year-olds were in first grade, and so on.

6) reported the very same inverted U-shaped evolution of Stroop interference, with a peak being located around grade 3.

Rosinski, Golinkoff, and Kukish (1975) exploited the picture-word variant of the Stroop task, in which interference on naming a picture is caused by an incongruent word superimposed on line drawing. They tested three groups of participants (grade 2, grade 4, and adults) and observed that picture-word interference decreased gradually with age.

To sum up, all the studies above concur to show a linear decrease in Stroop interference from 9-year olds onwards, but they differ for younger children. Some studies observed a decrease from the outset, while others reported an inverted U-shaped curve, with an initial increase in interference. Most of this apparent discrepancy certainly reflects the fact that the authors began their observations more or less early during development. Presumably, some investigators have not observed a first rising component all simply because the difficulty of carrying out the Stroop task in very young children dissuaded them from including these children in their observations.[16] The upward component of the curve, whether actually measured or not, is a logical requirement, because insofar as the effect exists in adult, it must emerge somewhere during reading acquisition. Consequently, the fact that the relation between Stroop interference and the amount of reading practice follows an inverted U-shaped function can reasonably be taken as a robust phenomenon.

A point of departure subsists, nevertheless, regarding the point where interference begins to decrease. It is possible that, for a given age group, the level of reading proficiency differ as a function of several factors, as suggested by the findings of Armengol (2002). Armengol tested first to sixth graders in order to provide normative developmental data on the Stroop effect in Spanish, using a procedure very similar to Comalli et al. (1962), except that she added a group of younger children. For children coming from private school, Armengol observed results very close to those of Comalli et al., namely a gradual decrease in Stroop interference from grade 1 to 6. For the children coming from public school, she replicated the inverted U-shaped curve obtained in many earlier studies, with an increase in Stroop interference from grade 1 to grade 2, followed by a gradual decrease. She explained the discrepancy between private and public schools for the first graders by a better reading proficiency at grade 1 in private school.

[16] Or at least in their report. Comalli et al. (1962) noted in their Footnote 4 that they also tested 5- and 6-year-old children, but they did not report the data.

Two contrasted interpretations

The literature reviewed up to now reveals two rather counterintuitive phenomena. The first is that a large effect emerges early during reading acquisition. The second counterintuitive phenomenon is the downward component of the inverted U-shaped curve. If Stroop interference reflected the automaticity of reading, and if the level of automaticity were a growing function of the amount of practice, then the capacity of reading to interfere with another less-practiced activity should appear gradually and should increase as reading skills improve. The observed pattern of performance suggests that the irrepressibility of automatic behavior does not grow as a monotonic function of practice, as do other characteristics of automaticity, such as speed and low attentional load. Supporting this view, Logan (1985) wrote that “skilled performers are usually able to control their performance better than unskilled performers, even though their performance is likely to be more automatic”, p. 379. Evidence for a better control has been observed by Logan (1982), who demonstrated that skilled typists were able to inhibit high-speed typing when detecting an error or an overt signal to stop. Tzelgov, Henik, and Leiser (1990), by testing bilinguals, observed that Stroop interference is controllable and that language proficiency is a precondition for such a control. Participants were able to reduce Stroop interference as a function of their expectancies in their native language but not in their second language. It is thus possible that the decrease in Stroop interference observed from childhood to adulthood reflects a genuine property of automatisms, whereby the possibility of cognitive control would increase with practice.

However, there is an alternative interpretation. In all the studies described above, the level of reading practice is entirely confounded with age, and it is possible that at least a part of the observed changes in Stroop interference is due to general factors evolving with age. Comalli et al. (1962) suggested that the amount of interference could be a positive function of the amount of practice, but that this effect would be overshadow by the age-related variations in the ability to inhibit nonrelevant information. Age-related variations in cognitive control would explain the decrease in Stroop interference observed from childhood to adulthood, because it is commonly acknowledged that this ability grows during the relevant period of time (e.g., Bédard et al., 2002; Klenberg, Korkman, & Lahti Nuuttila, 2001). This interpretation has been generally endorsed by the authors of subsequent studies.[17]

[17] This interpretation fits well with the observation that the amount of Stroop interference increases from young adults to older people (e.g., Comalli et al., 1962). Indeed, cognitive control could be less efficient in elderly than in young adults (Diamond, 2013). However, the increase of the Stroop effect is seemingly dependent on the version of the Stroop task (e.g., Ludwig, Borella, Tettamanti, & de Ribaupierre, 2010) and details of

Controlling for the influence of age

In principle, the solution for disentangling the competing views is straightforward: Because age appears as a confounding variable, the effect of reading practice should be assessed while keeping age constant. This is not possible, due to obvious practical and ethical constraints. A second-best solution consists in exploiting the natural variation in reading-skill abilities for a given age group. Fast readers may be considered as having more extensive reading practice, or at least as having better benefitted from practice, than slow readers. If the amount of Stroop interference were a positive function of practice, then fast readers should present greater interference than slow readers. Overall, empirical data confirm this prediction. In Martin (1978), adult participants were divided into two groups of fast and slow readers according to their mean reading speed of random word passages. Martin observed that fast readers showed significantly larger Stroop interference (181.51 ms) than slow readers (99.57 ms). Likewise, Stanovich, Cunningham, and West (1981) tested first graders with a variant of the Stroop task, in which children had to name the ink color of printed letters or noncolor words. At the end of the school year, the authors asked the teacher to classify the children according to their reading ability. A skilled group comprised the first 12 children on this scale and a less-skilled group comprised the last 12 children. Skilled readers consistently showed greater interference than less-skilled readers, although the differences were not significant. Moreover, all the children were administered several tests of reading (Reading Subtest Level I of the Wide Range Achievement Test, Reading subtest of the Standford Achievement Test, and reading a short paragraph orally that was timed). Correlational analyses showed a slight tendency for children who read the words faster and more accurately to exhibit greater interference.

More recently, Braet, Noppe, Wagemans, and Op de Beek (2011) investigated the relationship between Stroop interference and reading skills in the context of second-language acquisition in adults. First, they observed greater Stroop interference for native language (L1), for which reading is expected to be more automatic, than for second language (L2). Second, they found significant relations between Stroop interference and several measures of L2-reading skills, with greater interference linked to higher skills in L2.

All these studies are consistent with the idea that, when the influence of age is removed, the amount of Stroop interference could be positively related to reading practice. However, the strength of the evidence is weakened by the fact that individual variations in

measurement of Stroop interference (e.g., Verhaeghen & DeMeersman, 1998). This weakens any arguments based on the literature on aging.

reading skills for a given age group may depend on many cognitive and social factors in addition to the amount of reading practice. Moreover, the conclusions issued from studies comparing good and poor readers in the normal range are not confirmed by studies investigating reading disorders. In keeping with the data collected in the general population, people with reading disorders would be expected to exhibit a lesser level of Stroop interference than skilled readers. Paradoxically, studies exploiting this approach[18] have consistently observed more interference in dyslexics.

For instance, Everatt, Warner, Miles, and Thomson (1997) tested three groups of participants in the color-word Stroop task. The first group consisted of dyslexic children of 10 years old, the second group was a control group consisting of children of same chronological age, and the third group was composed of children matched with the reading age of the dyslexic children (chronological age: 8 years). The authors found that dyslexic children exhibited a level of Stroop interference that was similar to the interference observed with reading age-matched children, but *larger* than chronological age-matched children. More recently, Protopapas, Archonti, and Skaloumbakas (2007), Faccioli, Peru, Rubini, and Tassinari (2008) and Kapoula et al. (2010) replicated these findings. To account for their apparently paradoxical findings, the authors generally suggest that the stronger interference observed in dyslexics would not come from a potential impaired automaticity of reading in this population with regard to skilled readers, but rather would come from a reduced capacity to inhibit word reading.

To resume, studies investigating the evolution of Stroop interference with reading practice in normal children have reported an early emergence of a large Stroop effect, followed by a continuous decrement of the effect. The interpretation of these data is controversial, however, because reading ability is acquired within an age span where an overwhelming amount of other cognitive changes occurs, and the observed performance evolution may be due to multiple factors including a growing ability to inhibit irrelevant information. Studies controlling for age have relied on individual differences in reading abilities, and they report a larger Stroop interference in skilled readers than in less-skilled

[18] Hicks and Jackson (1981) assessed the development of the Stroop effect in dyslexics distributed as a function of their reading age (6-8, 8-10, 10-12, 12-14, and 14-16), but without a comparison with normal readers. The authors reported a decrease in Stroop interference as reading ages increased. However, participants in this study were between eight and 30 years of age, which entails that reading age at least partially coincided with chronological age (e.g., a 8-year-old dyslexic child necessarily falls into the 6-8 year reading age group). Given that developmental studies have consistently shown a decrease of the Stroop effect, at least for children above 7 to 9 years old, the decrease observed by Hicks and Jackson could be due to chronological age as well.

readers. However, again, the differences in reading abilities for a given age group are certainly related to a number of variables other than the amount of reading practice. Moreover, the comparison of skilled readers and dyslexics leads to the paradoxical conclusion that dyslexics exhibit a larger amount of Stroop interference.

The present study

Given that reading offers quite limited opportunity to investigate the effects of practice, we propose to investigate the question of how Stroop interference evolves with the amount of training through a paradigm that no longer relies on reading as the interfering process. To our knowledge, the only similar attempt was reported by MacLeod and Dunbar (1988). The authors used a Stroop-like task in which arbitrary associations were created across repeated experimental sessions, to ensure a better control on practice. Participants were trained to associate a color name to a shape during 20 days. In the experiment of concern here, participants were tested on different tasks on day 1, day 5, and day 20. The authors observed virtually no interference of shape on color naming on Day 1 (incongruent - control = 7 ms), while the interference became significant on day 5 (70 ms) and continued to increase on Day 20 (111 ms). To sum up, interference on color naming increased with shape-naming training. This result is consistent with the idea that Stroop interference would be positively related to training, but the question remains of whether a conclusion issued from a laboratory task, even extended along 20 sessions, can be generalized to the abilities resulting from the huge amount of practice common in everyday life.

In the present study, word reading, as source of interference, was traded against note naming in musicians. In earlier works, we have shown the existence of a Musical Stroop Effect (MSE) in adults (Grégoire, Perruchet, & Poulin-Charronnat, 2013, in press). The stimuli comprised a staff with a note in various positions. A name of a note was printed inside the note. In the congruent condition, the note name was congruent with the note position, while in the incongruent condition, name and position were incongruent. Musician participants were asked to read aloud the printed name of notes. The MSE designates the fact that reading note name was slowed down in the incongruent condition with regard to the congruent condition.

The musical Stroop paradigm shares most advantages of the classical color-word version to investigate automatisms. The most obvious is that note naming is an activity in which musician experts are intensively engaged over years, attaining a level of practice that is out of reach of laboratory settings. Moreover, note naming shares with word reading the nice property of being a component of more complex activities, the automaticity of which is required to ensure the successful expression of the whole behavior. The automaticity of note naming is necessary to allocate musicians' attention to higher integrative processes devoted to analyze chords and melodic lines of the musical work and to ensure motor control for instrument playing, as the automaticity of identification of single words is needed to free our minds to deal with higher-order aspects of reading such as comprehension and metacognitive functions. However, as an experimental means of investigation, the musical Stroop paradigm avoids the shortcomings pointed out above. In particular, for most children, musical training begins after they start to learn to read, making it possible to capture the effects of the early phase of practice while avoiding problems linked to testing very young children. Moreover, the inter-individual variations in the age to enter a music school makes it possible to decouple age and amount of musical practice, provided an appropriate selection of participants.

In Experiment 1, the development of the MSE was addressed by testing children in their first, second, and third year of music training. To anticipate, we did not observe the inverted U-shaped evolution in the amount of interference previously reported with reading in the standard color-word version of the task. However, it is possible that the amount of training was not sufficient to observe the downward component of the curve. In order to explore the evolution of interference along a larger amount of practice, Experiment 2 was run with children having received musical training up to five years.

Experiment 1

Method

Participants. Forty-two children from two music schools (the *Conservatoire à Rayonnement Régional* of Dijon and the École Municipale de Musique of Longvic) took part in the experiment. The children were divided into three groups (N = 14) according to their year of musical training; first, M_{age} = 7.51 years, SD = 0.27, 6 females; second, M_{age} = 8.65 years, SD = 0.47, 7 females; third, M_{age} = 9.36 years, SD = 0.72, 10 females. An additional group was made up of nonmusician children (N = 14, M_{age} = 7.70 years, SD = 0.46, 6 females). All the children were tested at the end of January. The children's vision was normal or corrected to normal and written parental consent was obtained for each child.

Material. The experimental material was composed of stimuli consisting of a treble staff with a note picture, which could appear on each of the 8 possible positions going from C4 to C5. The name of a note was written inside the note picture. For the congruent condition, the note name was congruent with the note position on the staff (Figure 1a), whereas in the incongruent condition, note name and position were incongruent, with the name written inside the note picture being one of the six other possible note names (e.g., when the note was DO, the written name was LA, SI, RE, MI, FA, or SOL; Figure 1b).

An additional test of note-naming abilities was designed to be closer to those involved in musical practice (Figure 1c). As for the congruent and the incongruent conditions, the note pictures from C4 to C5 were presented to the children, who had to name them.

To prevent the iconic memory of the staff from influencing the processing of the following note, the stimuli were randomly displayed at one of four possible positions without immediate repetition at the same location. The four positions were defined as the center of (invisible) rectangles resulting from the exhaustive partitioning of the screen into four quadrants of equal size. Stimuli were printed in black over a white background on a computer screen. Note names appeared in standard uppercase font 14. The treble staff was 7.7 cm wide by 5.1 cm high.

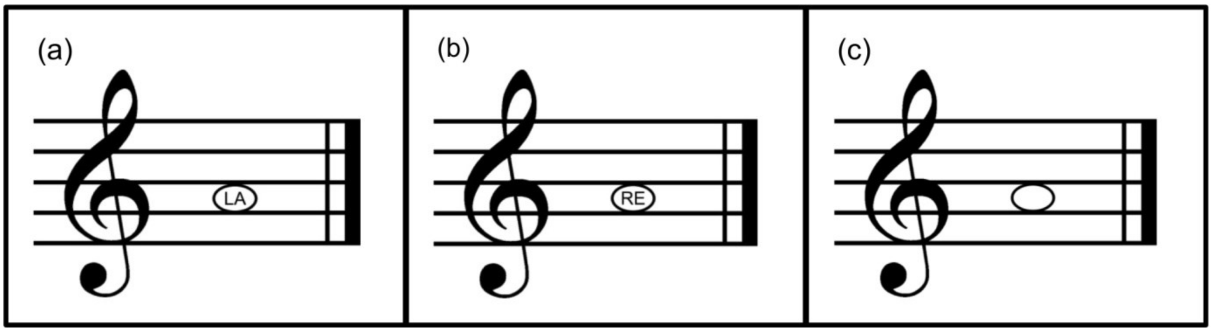

Figure 1. Examples of the different conditions used in Experiments 1 and 2: a) congruent condition; b) incongruent condition; c) note-naming ability test. Note that in the musical French notation (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

Procedure. The musician children had to perform a word-reading task, in which they had to read aloud the printed word while ignoring the note picture. There were two mixed conditions: congruent (Figure 1a) and incongruent (Figure 1b). For each condition, the stimuli appeared six times on each of the 8 locations, resulting in 48 trials per condition with a total of 96 trials (48 × 2). On each trial, a fixation cross was displayed for 1 s at the center of the screen before the occurrence of the stimulus, which stayed on the screen until participant's response. The interval between the response and the next trial was 1s. The trials were pseudo-randomly ordered for each child, excluding immediate repetitions of note locations or note names. They were displayed as four blocks of 24 trials each with a self-paced break between blocks.

The experimental session was immediately followed by an additional note-naming ability test, in which children had to name notes (Figure 1c). The ability test included 48 trials, the notes occurring 6 times each. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of notes. They were displayed as two blocks of 24 trials each with a self-paced break between blocks. The control nonmusician children only performed the word-reading task.

Children were encouraged to respond as fast and as accurately as possible throughout the session. The response times (RTs) were recorded by a voice key. During the session, the experimenter noted error responses and voice-key dysfunctions. After the experiment, the children filled out a questionnaire about their musical training.

Results

Ability test (musician children only)

Voice-key dysfunctions led to the exclusion of 13.89% of the data. The proportion of errors in the note-naming ability test was quite low, $M = 3.40\%$, $SD = 10.04$. Errors decreased with the year of musical training (Table 1). However, an ANOVA on note-naming errors performed with Year of Musical Training (first, second, and third year) as a between-subject variable showed no significant effect, $F(2, 39) = 1.00$, $p = .378$.

RTs for correct responses beyond three standard deviations of the mean (1.24%) were removed. The remaining data are presented in Figure 2, top panel. The mean RTs decreased with the year of musical training. An ANOVA revealed a significant main effect, $F(2, 39) = 5.61$, $p = .007$, $\eta_p^2 = .223$, with a significant linear trend, $p = .003$.

Note that there was no evidence for a speed-accuracy trade-off in the note-naming ability test, as shown by the positive correlation between errors and RTs, $r(40) = .868$, $p < .001$.

Word reading (MSE)

Voice-key dysfunctions led to the exclusion of 9.90% and 7.81% of the data for the musician and nonmusician children, respectively. A small proportion of errors was observed for the musician children (0.75%, *SD* = 2.09), while the nonmusician children made no errors of reading. An ANOVA on reading errors performed with Congruity (congruent, incongruent) as a within-subject variable and Year of Musical Training (nonmusicians, first, second, third) as a between-subject variable showed a significant main effect of congruity, $F(1, 52) = 4.60$, $p = .037$, $\eta_p^2 = .081$, with more errors in the incongruent condition than in the congruent condition. There was no effect of the year of musical training, $F(3, 52) = 1.95$, $p = .134$, and no Congruity × Year of Musical Training interaction, $F(3, 52) = 0.91$, $p = .440$.

Table 1
Proportions of errors as a function of years of musical training and condition in Experiment 1 and Experiment 2

Experiment 1				
	Nonmusicians	1 year	2 years	3 years
Ability test	None	5.97 (16.91)	3.59 (3.84)	0.63 (1.33)
Congruent	0 (0)	0.87 (1.61)	0.33 (1.24)	0 (0)
Incongruent	0 (0)	1.53 (1.68)	1.44 (4.24)	0.34 (0.88)

Experiment 2						
	Nonmusicians	1 year	2 years	3 years	4 years	5 years
Ability test	None	1.37 (2.29)	1.93 (5.36)	1.17 (2.12)	1.87 (2.74)	0 (0)
Congruent	0 (0)	0.40 (1.37)	0.79 (1.16)	0 (0)	0.39 (0.91)	0.25 (0.87)
Incongruent	0 (0)	1.21 (2.17)	1.64 (2.48)	0.59 (1.07)	1.55 (2.01)	1.38 (3.52)

RTs for correct responses beyond three standard deviations of the mean were removed, 0.67% and 1.29% for the musician and the nonmusician children, respectively. The remaining data are shown in Figure 2, bottom panel. An ANOVA on RTs carried out as for reading errors gave a significant effect of Year of Musical Training, $F(3, 52) = 8.63$, $p < .001$, $\eta_p^2 = .332$. The effect of congruity was also significant, $F(1, 52) = 20.29$, $p < .001$, $\eta_p^2 = .281$. The difference between incongruent and congruent trials increased numerically from the nonmusicians to the third year of musical training, but the Congruity × Year of Musical Training interaction did not reach significance, $F(3, 52) = 1.47$, $p = .234$. The nonmusician children and the first-year children showed no significant congruity effect, $t(13) = 1.53$, $p = .149$ and $t(13) = 0.98$, $p = .343$, respectively. In contrast, second- and third-year children showed a significant difference between congruent and incongruent conditions, $t(13) = 3.39$, $p = .005$, $d = 0.26$ and $t(13) = 3.14$, $p = .008$, $d = 0.22$, respectively.

Correlational analyses

The amplitude of the MSE was related to age, although the correlation was not significant, $r(40) = .211$, $p = .182$.[19] It is worth stressing that this correlation was positive, whereas the downward component of the inverted U-shaped relation described in the Introduction suggest a negative correlation between the amount of Stroop interference and age. A hypothesis is that the observed correlation could be mediated by the level of musical practice. Indeed, on average, older children had more musical practice. Partial correlations were carried out to address this issue. When controlling for Year of Musical Training, the partial correlation between MSE and age dropped to zero, $r(39) = .020$, $p = .902$, suggesting that performance was independent from age.

The amplitude of the MSE was also linked to the naming times in the note-naming ability test, $r(40) = -.300$, a correlation that just failed to reach significance, $p = .053$. When the level of practice was controlled, this correlation dropped but remained substantial, $r(39) = -.222$, $p = .163$. This suggests that the test of note-naming ability could capture some relevant individual differences, which operate in addition to the level of musical education. Note that, unsurprisingly, the naming times in the note-naming ability test significantly correlated with the year of musical training, but to an extent that left room to independent sources of influence, $r(40) = -.446$, $p = .003$.

[19] All the correlations were calculated on the groups of musicians only, because certain correlations involved the note-naming ability test, which nonmusicians did not undergo.

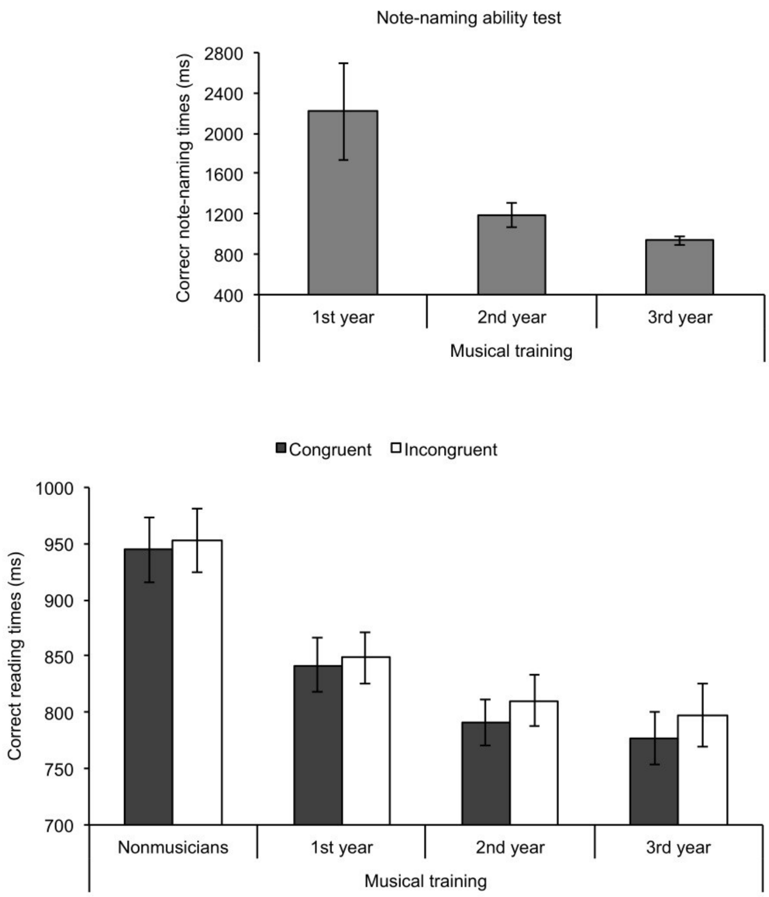

Figure 2. Correct note-naming times for the note-naming ability test (top panel), and correct reading times as a function of Congruity (congruent, incongruent; bottom panel) in Experiment 1. Both correct note-naming times and correct reading times are presented as a function of the Year of Musical Training (from nonmusicians to third year). Error bars indicate standard errors.

Discussion

The results of Experiment 1 confirms in children the results previously obtained in adults (Grégoire et al., 2013) regarding the existence of an MSE: On average, incongruent trials generated more errors and longer reading times than congruent trials. Finally, and more importantly, our study provides information about the evolution of the MSE with musical practice. Although the interaction between congruity and level of musical practice did not reach significance, there was a clear numerical trend for a continuous increase in the size of the MSE across practice.

A conservative conclusion is that this experiment does not replicate the inverted U-shaped curve that is commonly observed with the color-word Stroop task. This suggests that the downward part of the curve could be due to the fact that reading acquisition occurs during a life period where cognitive control increases. However, another possibility is that musical practice was not extensive enough to allow the observation of a decrement in the amplitude of the MSE. As outlined in the Introduction, studies carried out with the standard Stroop paradigm showed that the exact location of the peak of the inverted U-shaped curve varied from one study to another, and even if this location was precisely defined, generalization to the musical Stroop paradigm would be hazardous.

Experiment 2 aimed at addressing this concern, by exploring the effect of practice over five years of musical training instead of three. Even though the effect of age was found to be negligible in Experiment 1, a special effort was made in Experiment 2 to kept age as constant as possible across the five levels of musical training.

Experiment 2

Method

Participants. Sixty children from several music schools of Dijon and its suburbs took part in the experiment. None of them participated in Experiment 1. The children were divided into five groups (N = 12) according to their year of musical training; first, M_{age} = 9.56 years, SD = 0.74, 5 females; second, M_{age} = 9.09 years, SD = 0.84, 8 females; third, M_{age} = 10.32 years, SD = 0.56, 9 females; fourth, M_{age} = 10.58 years, SD = 0.43, 9 females; and fifth, M_{age} = 10.97 years, SD = 0.65, 8 females. A group made up of nonmusician children was added (N = 12, M_{age} = 9.76 years, SD = 0.21, 6 females). All the children were tested during the second semester. The children's vision was normal or corrected to normal and written parental consent was obtained for each child.

Material and Procedure. The material and the procedure were exactly the same as those in Experiment 1.

Results

Ability test (musician children only)

Voice-key dysfunctions led to the exclusion of 14.06% of the data. The proportion of errors in the note-naming ability test was quite low, M = 1.27%, SD = 3.01 (Table 1). An ANOVA on note-naming errors performed with Year of Musical Training (from the first to the fifth) as a between-subject variable showed no significant effect, $F(4, 55) = 0.79$, $p = .534$.

RTs for correct responses beyond three standard deviations of the mean (1.06%) were removed. The remaining data are shown in Figure 3, top panel. The mean RTs decreased with year of musical training. The main effect of the variable was significant, $F(4, 55) = 10.09$, $p < .001$, $\eta^2_p = .423$, with a significant linear trend, $p < .001$.

Note that there was no evidence for a speed-accuracy trade-off in the note-naming ability test, as shown by the positive correlation between errors and RTs, $r(58) = .229$, $p = .078$.

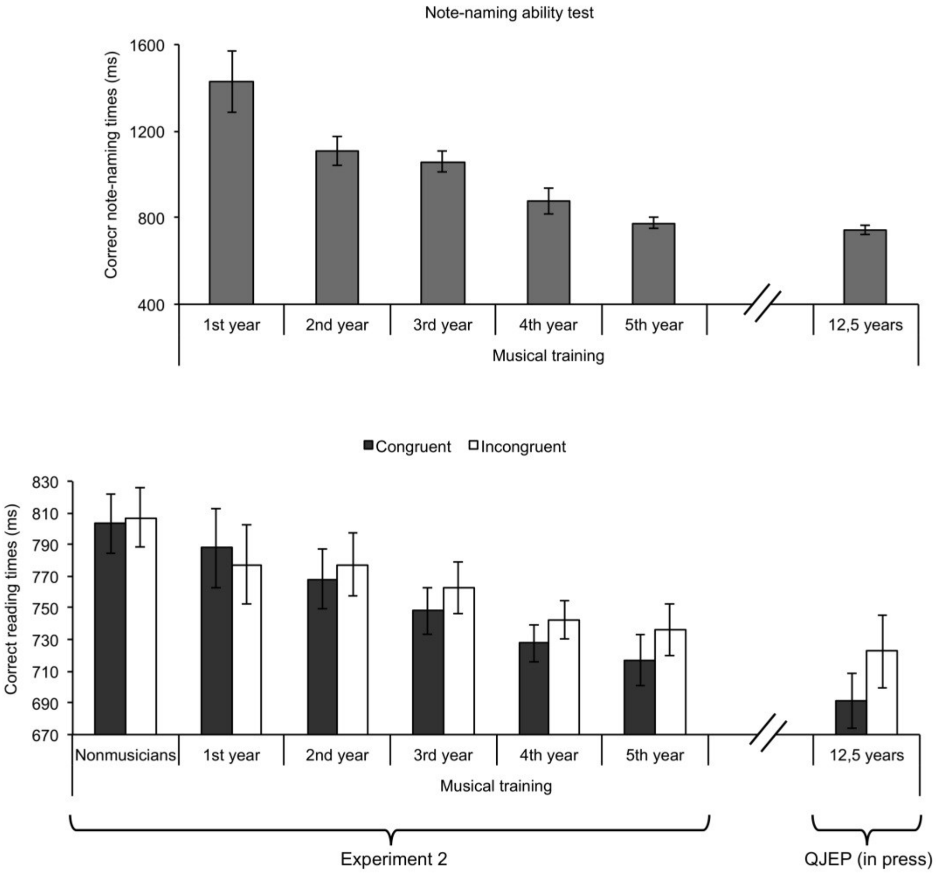

Figure 3. Correct note-naming times for the note-naming ability test (top panel), and correct reading times as a function of Congruity (congruent vs. incongruent; bottom panel) in Experiment 2. Both correct note-naming times and correct reading times are presented as a function of the Year of Musical Training (from nonmusicians to fifth year). The data collected by Grégoire et al. (in press) from young adult musicians (12,5 years of musical training) have been added for comparison (see General discussion). Error bars indicate standard errors.

Word reading (MSE)

Voice-key dysfunctions led to the exclusion of 9.98% and 6.34% of the data for the musician and the nonmusician children, respectively. A small proportion of errors was observed for the musician children (0.82%, *SD* = 1.84), while the nonmusician children made no errors of reading.

An ANOVA on reading errors performed with Congruity (congruent, incongruent) as a within-subject variable and Year of Musical Training (nonmusicians, first, second, third, fourth, and fifth) as a between-subject variable showed a main effect of congruity, $F(1, 55) = 7.85$, $p = .007$, $\eta_p^2 = .106$, with more errors in the incongruent condition than in the congruent condition. There was no effect of year of musical training, $F(5, 66) = 1.67$, $p = .153$, and no Congruity × Year of Musical Training interaction, $F(5, 66) = 0.42$, $p = .836$.

RTs for correct responses beyond three standard deviations of the mean were removed, 0.66% and 0.56% for the musician and the nonmusician children, respectively. The remaining data are shown in Figure 3, bottom panel. An ANOVA on RTs carried out as for errors revealed a main effect of the year of musical training, $F(5, 66) = 2.70$, $p = .028$, $\eta^2_p = .170$, with a significant linear trend, $p = .001$. There was also an effect of congruity, $F(1, 66) = 13.97$, $p < .001$, $\eta^2_p = .175$, which, importantly, was qualified by a significant Congruity × Year of Musical Training interaction, $F(5, 66) = 3.70$, $p = .005$, $\eta^2_p = .218$. For the nonmusician children and the first- and second-year children, there was no reliable difference between congruent and incongruent conditions, $t(11) = 0.61$, $p = .55$, $t(11) = 1.90$, $p = .083$ and $t(11) = 1.66$, $p = .125$, respectively. For the other children, the RTs for the incongruent condition were significantly longer than those of the congruent condition, $t(11) = 2.29$, $p = .043$, $d = 0.29$, $t(11) = 3.14$, $p = .009$, $d = 0.39$, and $t(11) = 3.64$, $p = .004$, $d = 0.37$, respectively for the third, fourth, and fifth year of musical training.

Correlational analyses

Although the age span of the children (1.5 year between the most extreme groups) was smaller than in Experiment 1, the amplitude of the MSE was again positively related to age, $r(58) = .249$, $p = .055$.[19] However, as in Experiment 1, when controlling for Year of Musical Training, the partial correlation dropped to zero, $r(57) = -.054$, $p = .686$, confirming that the MSE was independent from age.

The amplitude of the MSE was also significantly related to the naming times in the note-naming ability test, $r(58) = -.348$, $p = .006$. When the level of practice was controlled, this correlation dropped substantially, $r(57) = -.105$, $p = .427$. This suggests that the performance in the test of note-naming ability is more related to the level of musical training than in Experiment 1. Indeed, the naming times in the note-naming ability test correlated more strongly with the level of musical practice in Experiment 2, $r(58) = -.630$, $p < .001$, than in Experiment 1.

Discussion

The results of Experiment 2 confirmed and strengthened the main conclusions of Experiment 1. Incongruent trials generated more errors and longer reading times than congruent trials, providing an additional support for the existence of an MSE in musician children. The most important result, however, is related to the evolution of the MSE with musical practice. Experiment 1 concluded against the inverted U-shaped relation observed with the color-word version of the Stroop task, but evidence for a continuously growing MSE was limited for at least two reasons. First, the numerical increase failed to reach significance, and second, the observation concerned only the first three levels of musical training in music schools. Experiment 2 revealed a significant increase of the MSE with practice, which was extended to five years of musical training.

As a point of departure between the two experiments, a significant MSE appeared as soon as the second year of musical training in Experiment 1, whereas the effect became significant only from the third year in Experiment 2. This difference may be related to differences regarding the sampling of participants. In Experiment 1, most musician children came from the music conservatory of Dijon, the objective of which is to train professional musicians. In Experiment 2, the objective of reducing age differences between the five groups of musician children led to the recruitment being extended to a larger number of music schools from the suburban area of Dijon, making musical achievement more heterogeneous and certainly of a lower standard. This difference and its consequence is reminiscent of the study of Armengol (2002), who observed a one-year difference in the appearance of the classical Stroop effect between private and public schools.

General discussion

Only a very small proportion of the Stroop literature deals with the question of how Stroop interference evolves with the amount of reading practice. This is somewhat surprising in regard of the implications of the issue. As claimed by MacLeod (1991), "practice may turn out to be one of the most effective manipulations for disentangling theories of the Stroop effect" (p. 182). However, the few studies that have addressed the question convey a remarkably consistent picture. The relation between Stroop interference and the amount of reading practice takes the form of an inverted U-shaped curve, even though some studies may capture only a part of the whole shape. The Stroop effect appears early during reading acquisition, reaches a peak around grade 3, then continuously decreases until adulthood (Armengol, 2002; Comalli et al., 1962; Dash & Dash, 1982; Peru et al., 2006; Rand et al., 1963; Rosinski et al., 1975; Schadler & Thissen, 1981; Schiller, 1966).

The downward component of the curve is rather counterintuitive, because a natural view is that irrepressibility should increase continuously with practice, or alternatively should reach an asymptote after some amount of practice. The observed decrease makes sense if one considers that skilled performers are able to control their performance better than less-skilled performers, a view that has been advocated by a few authors (Logan, 1985; Tzelgov et al., 1990). However, it is worth stressing that in all relevant studies, the level of reading practice was entirely confounded with age due to obvious practical and ethical constraints, and it is possible that at least a part of the observed changes in Stroop interference was due to general factors evolving with age. In particular, the decrease in Stroop interference observed after grades 2 or 3 could be due to a growing ability to inhibit nonrelevant information (Diamond, 2013). In partial support for this view, studies relying on individual differences in reading abilities for a given age group report a larger Stroop interference in skilled readers than in less-skilled readers (Braet et al., 2011; Martin, 1978; Stanovich et al., 1981).

Summary of the main results

In the present study, word reading, as source of interference, was traded against note naming in musicians. Grégoire et al. (2013, in press) demonstrated the existence of an MSE in adults, whereby reading a note name written inside a note on a staff was slower when note name and note location were incongruent than when note name and note location were congruent. Like word reading, note naming is an activity that is intensively practiced over years by musician experts but, in contrast to word reading, the level of practice on note naming does not necessarily coincide with the different grades of elementary school. In particular, the two reported experiments did not involve the very young children who are necessarily included in the study population whenever one intends to track the early emergence of the standard Stroop effect. Moreover, in Experiment 2, in which all children were over 9 years of age, the age difference between the extreme groups (first vs. fifth year of musical training) was reduced to 1.5 year, which is obviously impossible when reading is the interfering process.

In these conditions, we also observed an early appearance of the interference effect, as reported for the color-word Stroop task, but we did not replicate the inverted U-shaped curve. Both Experiments 1 and 2 revealed a linear and positive relation between the amplitude of the MSE and the amount of musical practice, which reached significance when 5 years of musical training were considered in Experiment 2. Note that this gradual increase of the MSE is all the more noticeable as it occurred within a context where the overall reading times substantially decreased. This means that the effect cannot be accounted for by some kind of floor effect, whereby shorter RTs would offer smaller space for an effect to emerge. If assessed as a proportion of the overall reading times instead of as a simple arithmetic difference between incongruent and congruent trials, the gradual increase of the MSE would be still more salient.

Could additional training reveal an inverted U-shaped relation?

Are five years of training sufficient to definitely rule out the existence of a downward component to the relation between the MSE and musical practice? It is clear that all Stroop studies described above report a decrement of the Stroop interference long before five years of reading practice. It could be argued that using the number of years of practice as a benchmark is misguiding, because even for musicians, note naming is certainly a less pervasive activity than reading. Accordingly, collecting the actual amount of weekly practice in note-naming and reading tasks would be more appropriate. More importantly, however, directly comparing the duration of practice between two tasks, whatever the measure, makes little sense. Indeed, the level of automaticity resulting from a same amount of practice obviously depends on the processes involved in the tasks. For instance, Logan and Klapp (1991) have shown that genuine properties of automaticity can emerge after less than 15 min of training on a simple and consistent alphabet-arithmetic task, an amount of practice that is not in the order of magnitude required to master complex activities such as reading. For our concern, the musical notation looks as much simpler than the writing system, running counter the hypothesis that a decrement of the MSE could appear after 5 years of musical training.

In order to definitely rule out this hypothesis however, is it possible to explore the evolution of the MSE over a longer time period while keeping age constant? Whereas note naming allows a much better control to tease apart age and practice than word reading, the task does not release from any practical constraints nevertheless. For instance, to extend the investigation to ten years of practice, one would need to recruit, say, a group of 16-year-old young adults who began music learning when they were 6 years old (which is manageable), but also, to consider only the other end on the level of practice, a number of participants of the same age who are just beginning to learn music. This may turn out to be a major challenge. However, it must be realized that keeping age constant may be no longer necessary at this stage. First, the capacity of cognitive control of 10-11 years old children (the age range of participants in Experiment 2) is certainly close to their adult achievement, even though further improvement cannot be excluded (e.g., Diamond, 2013). Second and more importantly, the effects of the residual differences in cognitive control are not necessarily a problem. Indeed, the better control of adults can only result in a decrease of the effect. As a consequence, if the MSE turns out to decrease between 10-11 years olds and adults, the effect of age will remain impossible to disentangle from the effect of practice. But if an MSE increment is observed, this change ought to be attributed to adults' extended practice. To address this issue, the data from Experiment 2 were compared to data collected in young adult musicians whose the mean amount of musical practice was 12.5 (SD = 4.51) years (Grégoire et al., in press).[20] The results are shown in Figure 3. It appears that adult performances provide a nearly perfect continuation of children curves: The reading times for both congruent and incongruent stimuli still decrease, and crucially, the difference between the two values still increases. The MSE reached 31.36 ms in adults, whereas the corresponding value for children after five years of musical practice was 18.91 ms. These data make it unlikely that a too short amount of practice would be responsible for the observation of a monotonic increase of the MSE with musical practice.

About generalization

The strength of the evidence for a monotonic relation between practice and Stroop interference for note naming leaves open the question of whether the same conclusion holds for other tasks, and first of all, for reading. The Stroop literature includes several studies that made use of other tasks than reading as the interfering process to address specific issues. For instance, Palef and Olson (1975) exploited the automatic processing of spatial position, MacLeod and Dunbar (1988) investigated the automaticity of an arbitrary association learned in laboratory setting, and Akiva-Kabiri and Henik (2012) focused on the automaticity of pitch coding in absolute pitch possessors. By and large, the fact that such studies are informative about general properties of automaticity, even though reading is not the interfering process, is at least tacitly acknowledged, suggesting that our conclusions could be also applied to other tasks. However, a contradictor could object that our situation embeds a particular feature, which could preclude the generalization of our results: In our experiments, the time required to name the note (the task-irrelevant dimension) was longer than the time required for word reading (the task-relevant dimension), even after 5 years of musical practice. This relation is

[20] We choose these data as a point of comparison, because experimental conditions were closer to the present ones than in Grégoire et al. (2013). The conditions were not exactly identical, but the differences went in a direction that would tend to reduce the magnitude of the MSE in adults (in particular, children were tested with the most familiar octave, C3-C4, whereas adults were tested with two octaves).

in the opposite direction to the relation observed in the standard color-word or picture-word situations, and is reminiscent of what is usually coined as the Reverse Stroop Effect (RSE). In studies on the RSE, the question is whether *reading* words may be impaired by an incongruent color-word or picture-word combination. The overall picture is that the RSE is, at best, an elusive phenomenon, and in this regard, the RSE seems to differ strikingly from the original Stroop effect (for a review: Grégoire et al., in press).

We believe that this line of reasoning amounts to mistake a statistically prevalent characteristic of the Stroop task for a theoretical prerequisite. It has been shown for long that the most common pattern of latencies—the latencies of responses being shorter for the irrelevant than for the relevant dimension—is not a criterion defining a Stroop effect. For instance, in Dunbar and MacLeod (1984), reading was made difficult and the authors wrote in their abstract: "even when reading a color word was considerably slower than naming the color of ink in which the word was printed, Stroop interference persisted virtually unaltered". The authors did not consider that, in this case, the effect differs in nature and should be renamed "reverse Stroop". The likelihood that our conclusion on the monotonic relation between practice and the size of the Stroop effect also applies to other automatic tasks, and in particular to reading, is further strengthen by (1) the converging evidence collected in fast and slow readers, with fast readers exhibiting a stronger Stroop effect than slow readers, and (2) the fact that data running against this relation—the downward component of the acquisition curves and the data collected in dyslexics—may be easily explained in terms of cognitive control. Admittedly, a skeptical position cannot be definitely ruled out, but given that our experimental approach cannot be transposed in a reading task, such a position does not meet the criterion of falsifiability.

To sum up, a reasonable conclusion is that the inverted U-shaped relation between Stroop interference and reading practice that has been reported up to now seems to be linked to the fact that reading practice largely coincides with age, and that the downward component of the curve depends on the growing children's ability to inhibit irrelevant information. When age no longer coincides with practice, irrepressibility seems to be a monotonic function of practice. There is a long-standing debate regarding the question of whether the different properties commonly attributed to automaticity evolve in parallel (e.g., Johnston & Dark, 1986; Paap & Odgen, 1981). Regarding the speed of processing and the independence from attentional resources, there is a consensus that more practice leads to a gradual and continuous improvement. Earlier results on the evolution of the Stroop effect could have suggested that irrepressibility has to be dissociated from other properties. Our results suggest that such a conclusion could be unwarranted, and that the amount of interference is a monotonic function of practice, as presumably the other characteristics of automaticity.

References

Akiva-Kabiri, L. & Henik, A. (2012). A unique asymmetrical stroop effect in absolute pitch possessors. *Experimental Psychology, 59*(5), 272-278.

Armengol, C. G. (2002). Stroop test in Spanish: Children's norms. *Clinical Neuropsychologist, 16*, 67-80.

Bédard, A. C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology, 21*, 93-111.

Braet, W., Noppe, N., Wagemans, J., & Op de Beeck, H. (2011). Increased Stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology, 64*, 596-607.

Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interfernce effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology, 100*, 47-53.

Dash, J., & Dash, S. (1982). Cognitive developmental studies of the Stroop phenomena: Cross-sectional and longitudinal data. *Indian Psychologist, 1*, 24-33.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 622-639.

Everatt, J., Warner, J., Miles, T. R., & Thomson, M. E. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia, 3*, 222-228.

Faccioli, C., Peru, A., Rubini, E., & Tassinari, G. (2008). Poor readers but compelled to read: Stroop effects in developmental dyslexia. *Child Neuropsychology, 14*, 277-283.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269-278.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (in press). About the unidirectionality of interference: Insight from the Musical Stroop Effect. *The Quarterly Journal of Experimental Psychology*.

Hicks, C., & Jackson, P. (1981). A study of the relationship between the Stroop effect and reading age in dyslexic subjects. *Journal of Research in Reading, 4*, 29-33.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology*, 37, 43-75.

Kapoula, Z., Le, T. T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., et al. (2010). Poor Stroop performances in 15-year-old dyslexic teenagers. *Experimental Brain Research, 203*, 419-425.

Klenberg, L., Korkman, M., & Lahti Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20*, 407-428.

Logan, G. D. (1982). On the ability to inhibit complex movements: A Stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 778-792.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39*, 367-386.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is Extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 179-195.

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the Color Stroop Test: A comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics, 51*(2), 135-142.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126-135.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*, 383-391.

Martin, M. (1978). Speech recoding in silent reading. *Memory & Cognition, 6*, 108-114.

Paap, K. R., & Ogden, W. C. (1981). Letter encoding is an obligatory but capacity-demanding operation. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 518-527.

Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology, 29*(3), 201-209.

Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes de Biologie, 144*, 45-62.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology, 54*, 251-282.

Rand, G., Wapner, S., Werner, H., & McFarland, J. H. (1963). Age differences in performance on the Stroop color-word test. *Journal of Personality, 31*, 534-558.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46*, 247-253.

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9*, 132-141.

Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology, 72*, 105-108.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior, 13*, 57-74.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a biligual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 760-771.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging, 13*(1), 120-126.

Etude 3b

Introduction

L'objectif de cette étude était d'examiner, de manière plus fine que dans l'Etude 3a, l'évolution du caractère obligatoire de la dénomination de notes chez des enfants musiciens en début d'apprentissage du solfège. De nombreux résultats expérimentaux indiquent que l'interférence, dans une tâche de Stroop couleur-mot, augmente fortement en début d'apprentissage de la lecture (e.g., Ehri, 1976; Stanovich, Cunningham, & West, 1981) puis diminue ensuite avec un surcroît de pratique (e.g., Ehri & Wilce, 1979). Nous souhaitions vérifier si l'augmentation obtenue entre le MSE et le niveau de pratique du traitement non-pertinent (i.e., la dénomination de notes), dans l'Expérience 1 de l'Etude 3a[21], ne dissimulait pas une forte croissance suivie d'une diminution entre les deux premiers groupes testés. Une augmentation plus modérée apparaîtrait alors par la suite à partir de la deuxième année de solfège. Bien que ce pattern de résultat puisse sembler improbable, il a déjà été mis en évidence par Chen et Ho (1986) avec des sujets bilingues. En utilisant le paradigme de Stroop avec des mots appartenant à une seconde langue, ces auteurs ont constaté que les participants (des chinois qui apprenaient l'anglais) avaient une interférence importante en début d'apprentissage, qui diminuait quand le niveau de pratique (de l'anglais) s'accentuait, puis augmentait à nouveau chez les sujets les plus expérimentés.

Pour tester cette hypothèse, nous avons réalisé trois passations durant la même année d'enseignement avec trois groupes de sujets (en première, deuxième et troisième année de solfège). Nous avons ainsi obtenu une mesure détaillée de l'évolution du MSE en fonction du niveau de pratique musicale en début d'apprentissage du solfège.

Méthode

Les participants, le matériel et la procédure étaient ceux de l'Expérience 1 de l'Etude 3a. La première passation correspond aux données présentées dans l'Expérience 1 de l'Etude 3a (au mois de janvier), puis deux autres passations ont été réalisées (aux mois de mars et mai de la même année) pour tous les sujets.

[21] Cette étude est un complément de l'Expérience 1 de l'Etude 3a. Les participants de l'Expérience 1 avaient en fait été testés trois fois au cours de l'année scolaire. Les données des deux dernières passations n'ont pas été intégrées dans l'Etude 3a car les résultats n'apportent pas d'informations claires et convaincantes sur l'évolution de l'automaticité, comme nous le détaillerons par la suite.

Résultats

Le test d'habiletés en dénomination de notes

Les dysfonctionnements de la clé vocale ont conduit à exclure 9.59% des données. Une ANOVA réalisée sur les proportions d'erreurs avec le facteur Passation (première, deuxième, troisième) comme variable intra-sujets et le facteur Groupe (1, 2, 3) comme variable inter-sujets indiquait un effet principal tendanciel de la Passation, $F(2, 78) = 2.98$, $p = .057$, $\eta^2_p = .071$, pas d'effet principal du Groupe, $F(2, 39) = 1.30$, $p = .283$, ni d'interaction Passation × Groupe, $F(4, 78) = 0.76$, $p = .557$. Les proportions d'erreurs augmentaient entre la première et la troisième passation pour le groupe 1 (5.97%, 1.92%, et 8.51%, respectivement), comme pour le groupe 3 (0.63%, 0.66%, et 2.64%, respectivement). Les tendances linéaires étaient significatives, $F(1, 39) = 8.76$, $p = .005$, $\eta^2_p = .183$, et $F(1, 39) = 5.55$, $p = .024$, $\eta^2_p = .125$, respectivement. Ces résultats sont intrigants car ils semblent indiquer que les sujets des groupes 1 et 3 sont devenus moins performants en dénomination de notes entre la première et la troisième passation, alors qu'un résultat inverse était attendu. En revanche, les pourcentages d'erreurs n'augmentaient pas significativement en fonction des passations pour le groupe 2 (3.59%, 1.79%, et 3.53%, respectivement), $F(1, 39) = 0.01$, $p = .940$.

Les TRs des réponses correctes au-delà de trois écarts-types de la moyenne (1.47%) ont été retirés. Les données restantes sont représentées sur la Figure 1. Une ANOVA sur les TRs réalisée avec le facteur Passation (première, deuxième, troisième) comme variable intra-sujets et le facteur Groupe (1, 2, 3) comme variable inter-sujets révélait un effet principal significatif du Groupe, $F(2, 39) = 9.68$, $p < .001$, $\eta^2_p = .332$, mais pas d'effet principal de la Passation (première, deuxième, troisième), $F(2, 78) = 1.97$, $p = .146$, ni d'interaction Passation × Groupe, $F(4, 78) = 1.51$, $p = .206$. Les TRs diminuaient progressivement en fonction des passations pour les enfants du premier groupe. La tendance linéaire était significative, $F(1, 39) = 6.24$, $p = .017$, $\eta^2_p = .138$. Aucune tendance linéaire ne fut observée dans les groupes 2 et 3, $F(1, 39) = 0.08$, $p = .781$, $F(1, 39) = 0.02$, $p = .899$, respectivement.

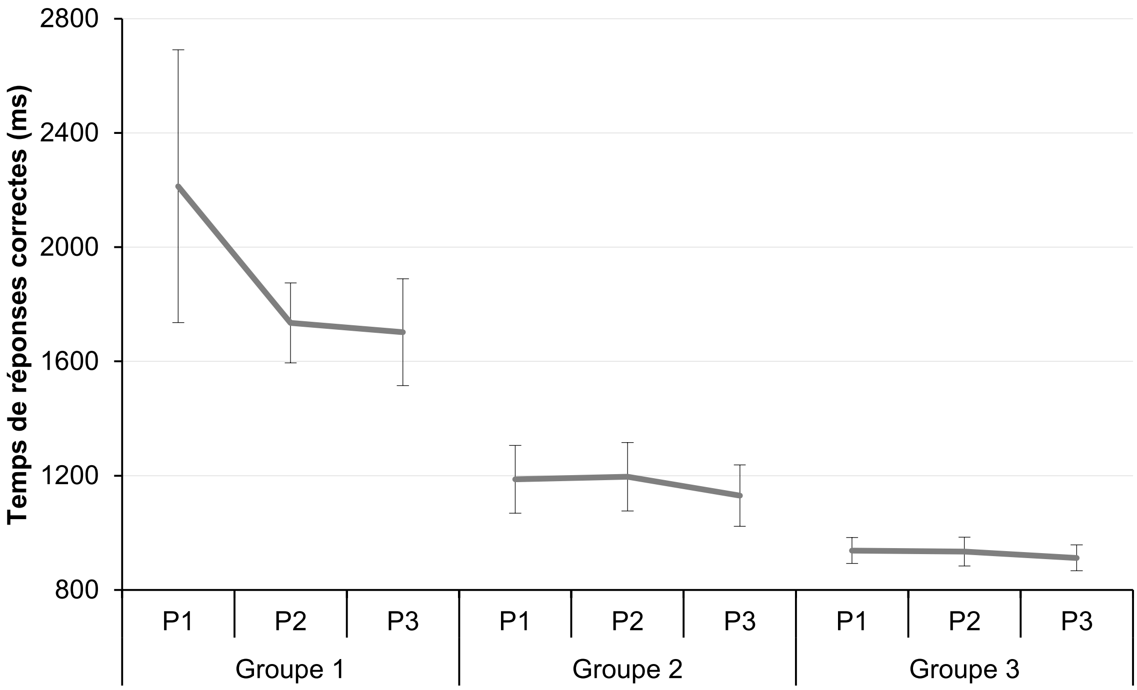

Figure 1. Temps de réponses correctes en fonction du Groupe et de la passation (P) dans le test d'habiletés en dénomination de notes. Les barres d'erreurs indiquent les erreurs standards.

La tâche de Stroop musical

Analyse des proportions d'erreurs

Les dysfonctionnements de la clé vocale ont conduit à exclure 7.62% des données. Une ANOVA réalisée sur les proportions d'erreurs avec le facteur Passation (première, deuxième, troisième) et le facteur Congruité (congruente, incongruente) comme variables intra-sujets et le facteur Groupe (1, 2, 3) comme variable inter-sujets révélait un effet principal de la Congruité, $F(1, 39) = 9.10$, $p = .004$, $\eta^2_p = .189$, mais aucun effet principal du Groupe, $F(2, 39) = 159$, $p = .854$, ni de la Passation, $F(2, 78) = 0.66$, $p = .522$. Aucun effet d'interaction n'atteignait le seuil de significativité : Passation × Groupe, $F(4, 78) = 1.46$, $p = .222$, Congruité × Groupe, $F(2, 39) = 0.67$, $p = .520$, Passation × Congruité, $F(2, 78) = .17$, $p = .844$, Passation × Congruité × Groupe, $F(4, 78) = 1.38$, $p = .247$.

Analyse des proportions d'erreurs pour le groupe 1. La proportion d'erreurs n'était pas significativement plus importante dans la condition incongruente que dans la condition congruente pour la passation 1, 1.53% vs. 0.87%, $t(13) = 1.28$, $p = .224$, la passation 2, 0.63% vs. 0.31%, $t(13) = 1.01$, $p = .330$, et la passation 3, 0.78% vs. 0.67%, $t(13) = 0.24$, $p = .818$. L'effet de la congruité, évalué par la différence entre la condition incongruente et la condition congruente, n'évoluait pas significativement en fonction des passations. Aucune tendance, ni linéaire, $F(1,39) = 0.39$, $p = .534$, ni quadratique, $F(1,39) = 0.02$, $p = .889$, ne fut observée.

Analyse des proportions d'erreurs pour le groupe 2. Les résultats sont similaires à ceux obtenus pour le groupe 1. La proportion d'erreurs n'était pas significativement plus importante dans la condition incongruente que dans la condition congruente pour la passation 1, 1.44% vs. 0.33%, $t(13) = 1.38$, $p = .192$, la passation 2, 0.65% vs. 0.17%, $t(13) = 1.33$, $p = .208$, et la passation 3, 1.03% vs. 0.62%, $t(13) = 0.68$, $p = .511$. L'effet de la congruité n'évoluait pas significativement en fonction des passations. Aucune tendance, ni linéaire, $F(1,39) = 0.60$, $p = .444$, ni quadratique, $F(1,39) = 0.38$, $p = .541$, ne fut observée.

Analyse des proportions d'erreurs pour le groupe 3. La proportion d'erreurs n'était pas significativement plus importante dans la condition incongruente que dans la condition congruente pour la passation 1, 0.34% vs. 0%, $t(13) = 1.47$, $p = .165$, et la passation 2, 1.15% vs. 0.32%, $t(13) = 1.23$, $p = .239$, mais un effet tendanciel fut obtenu pour la passation 3, 1.83% vs. 0%, $t(13) = 2.10$, $p = .056$, $d = 0.855$. L'effet de la congruité n'évoluait pas significativement en fonction des passations. Aucune tendance, ni linéaire, $F(1,39) = 2.75$, $p = .105$, ni quadratique, $F(1,39) = 0.32$, $p = .576$, ne fut observée.

Analyse des TRs

Les TRs des réponses correctes au-delà de trois écarts-types de la moyenne (0.79%) ont été retirés. Les données restantes sont représentées sur la Figure 2. Une ANOVA réalisée sur les TRs avec le facteur Passation (première, deuxième, troisième) et le facteur Congruité (congruente, incongruente) comme variables intra-sujets et le facteur Groupe (1, 2, 3) comme variable inter-sujets révélait un effet principal de la Congruité, $F(1, 39) = 22.62$, $p < .001$, $\eta_p^2 = .367$, un effet principal tendanciel du Groupe, $F(2, 39) = 2.75$, $p = .076$, mais aucun effet de la Passation, $F(2, 78) = 0.05$, $p = .949$. Aucun effet d'interaction n'atteignait le seuil de significativité : Passation × Groupe, $F(4, 78) = 0.57$, $p = .685$, Congruité × Groupe, $F(2, 39) = 0.77$, $p = .469$, Passation × Congruité, $F(2, 78) = 1.45$, $p = .241$, Passation × Congruité × Groupe, $F(4, 78) = 0.54$, $p = .710$.

Analyse des TRs pour le groupe 1. Les TRs n'étaient pas significativement plus importants dans la condition incongruente que dans la condition congruente pour la passation 1, 848.50 ms vs. 842.11 ms, $t(13) = 0.98$, $p = .343$, la passation 2, 857.93 ms vs. 849.66 ms, $t(13) = 1.25$, $p = .234$, et la passation 3, 862.93 ms vs. 856.40, $t(13) = 1.76$, $p = .102$. L'effet de la Congruité n'évoluait pas significativement en fonction des passations. Aucune tendance, ni linéaire, $F(1,39) = 0.00$, $p = .986$, ni quadratique, $F(1,39) = 0.09$, $p = .767$, ne fut observée.

Analyse des TRs pour le groupe 2. Les TRs étaient significativement plus importants dans la condition incongruente que dans la condition congruente pour la passation 1, 810.25 ms vs.

790.85 ms, $t(13) = 3.39$, $p = .005$, $d = 0.256$, mais pas pour la passation 2, 802.89 ms vs. 794.33 ms, $t(13) = 1.55$, $p = .146$, ni pour la passation 3, 806.49 ms vs. 795.66, $t(13) = 1.73$, $p = .107$. L'effet de la Congruité n'évoluait pas significativement en fonction des passations. Aucune tendance, ni linéaire, $F(1,39) = 1.29$, $p = .263$, ni quadratique, $F(1,39) = 1.16$, $p = .287$, ne fut observée.

Analyse des TRs pour le groupe 3. Les TRs étaient significativement plus importants dans la condition incongruente que dans la condition congruente pour la passation 1, 797.62 ms vs. 777.47 ms, $t(13) = 3.14$, $p = .008$, $d = 0.223$, de même que pour la passation 2, 793.92 ms vs. 782.97 ms, $t(13) = 2.18$, $p = .049$, $d = 0.128$, mais pas pour la passation 3, 779.45 ms vs. 769.88 ms, $t(13) = 1.52$, $p = .152$. L'effet de la Congruité n'évoluait pas significativement en fonction des passations. Aucune tendance, ni linéaire, $F(1,39) = 1.97$, $p = .168$, ni quadratique, $F(1,39) = 0.42$, $p = .523$, ne fut observée.

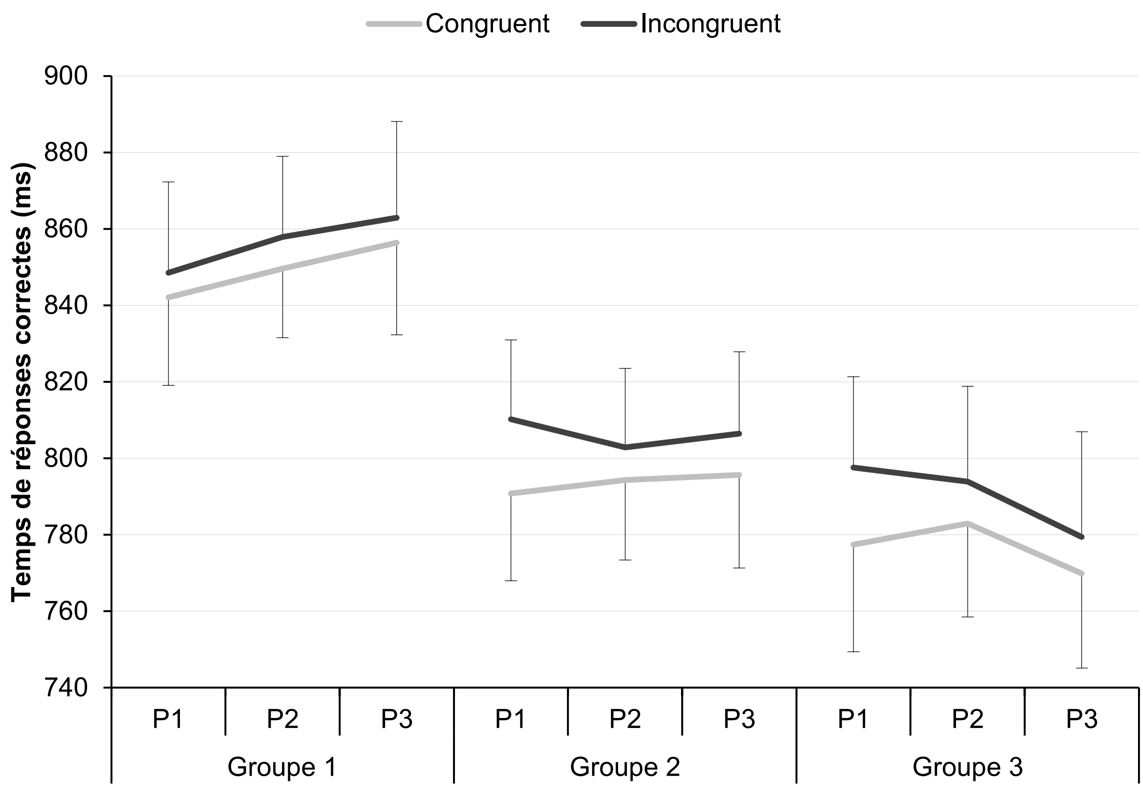

Figure 2. Temps de réponses correctes en fonction du Groupe, de la Congruité, et de la passation (P). Les barres d'erreurs indiquent les erreurs standards.

Discussion

L'objectif de cette étude était de réaliser une mesure fine de l'évolution du MSE en fonction du niveau de pratique musicale au cours des trois premières années de solfège. Nous avons effectué trois passations pendant la même année d'enseignement avec trois groupes de participants (en première, deuxième et troisième année de solfège). Les résultats indiquaient que le MSE n'apparaissait pas significativement au cours de la première année de solfège pour chacune des trois passations. L'effet de la congruité n'évoluait pas de façon significative en fonction du niveau de pratique de la dimension interférente (i.e., la dénomination de notes). Même à un niveau descriptif, l'amplitude de l'effet restait relativement stable : 6.39 ms pour la première passation, 10.95 ms pour la deuxième, et 6.53 ms pour la dernière. Les enfants en deuxième année de solfège montraient un MSE lors de la première passation, mais curieusement, cet effet disparaissait lors des deux passations suivantes (l'amplitude de l'effet s'atténuait et n'atteignait plus le seuil de significativité). De manière analogue, le MSE était présent chez les sujets du troisième groupe lors de la première passation, diminuait lors de la seconde passation (mais restait significatif), et n'apparaissait plus au moment de la troisième passation. L'atténuation et la disparition du MSE dans les groupes 2 et 3 semblent relativement énigmatiques, de même que l'absence d'augmentation de l'effet de la congruité à la fin de la première année de solfège (pour le groupe 1). Il paraît difficile d'expliquer pourquoi le MSE disparaissait à la fois dans les groupes 2 et 3, et pourquoi il n'apparaissait pas dans le groupe 1, malgré une plus grande pratique du traitement interférent.

Nous pourrions supposer que les enfants étaient moins concentrés dans la réalisation de leurs tâches lors de la deuxième et de la troisième passation, car ils connaissaient déjà la situation expérimentale. Néanmoins, les sujets n'ont semblé montrer aucun signe de relâchement, et ils étaient encouragés à rester attentifs à plusieurs reprises au cours de chaque passation. De plus, les résultats obtenus dans le test d'habiletés en dénomination de notes paraissent contredire, en partie, cette hypothèse. En effet, les TRs avaient tendance à diminuer dans cette tâche, pour les trois groupes de participants, en fonction des passations (même si cette diminution n'était pas significative pour les groupes 2 et 3). Cependant, l'augmentation de la proportion d'erreurs pour les groupes 1 et 3 rend les interprétations difficiles.

Il semblerait toutefois plus probable que la répétition de la tâche de Stroop musical mène les sujets à être moins sensibles au traitement non-pertinent. En effet, il a été montré que la pratique répétée de la situation Stroop classique conduisait à réduire l'interférence (e.g., Ogura, 1980). Ainsi, la pratique répétée du paradigme de Stroop musical masquerait les effets dus à une plus grande expertise de la dénomination de notes, ce qui expliquerait la réduction du MSE observée dans les groupes 2 et 3, de même que l'absence d'augmentation de l'effet de la congruité dans le groupe 1. Il nous paraît donc inapproprié d'étudier l'évolution du MSE de manière longitudinale sur des délais relativement courts (environ 8 à 10 semaines entre deux passations dans cette étude). C'est pourquoi, nous privilégierons par la suite une approche transversale.

Bibliographie

Chen, H. C, & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 397–401.

Ehri, L. C. (1976). Do words really interfere in naming pictures? *Child Development, 47*, 502–505.

Ehri, L. C., & Wilce, L. S. (1979). Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology, 27*, 352–364.

Ogura, C. (1980). Formation of learning-set in the task of Stroop color-word test. *Journal of Child Development, 16*, 30-36.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior, 13*, 57–74.

Etude 4

How are two automatisms developing concurrently with practice?
Pitting note naming against reading in musicians

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat
CNRS-UMR 5022, University of Bourgogne, Dijon, France

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (in preparation). How are two automatisms developing concurrently with practice? Pitting note naming against reading in musicians.

Author Note

Laurent Grégoire, Pierre Perruchet, and Bénédicte Poulin-Charronnat, Université de Bourgogne, LEAD/CNRS, UMR5022, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France

This research was supported by the grant ANR-09-JCJC-0129 and the grant Conseil Régional de Bourgogne N° 2012-9201AAO05S01353. The authors are grateful to Cédric Foucault for help with collecting the data.

Correspondence concerning this article should be addressed to Laurent Grégoire, LEAD-CNRS UMR5022, Université de Bourgogne, Pôle AAFE, 11 Esplanade Erasme, 2100 Dijon, France. E-mail: laurent.gregoire@u-bourgogne.fr

Résumé

Dans les deux tâches du paradigme de Stroop couleur-mot, une asymétrie est habituellement observée, avec une interférence beaucoup plus importante de la lecture sur la dénomination de la couleur (effet Stroop) que vice versa (effet Stroop inversé). Quand un effet Stroop inversé est obtenu, c'est la plupart du temps aux dépens de l'effet Stroop lui-même, comme si ces deux effets ne pouvaient pas coexister. En utilisant le paradigme de Stroop musical, nous avons observé que les participants montraient un effet Stroop musical (MSE) et un effet Stroop musical inversé (RMSE; Grégoire, Perruchet, & Poulin-Charronnat, sous presse). La présente étude évalue l'évolution conjointe des deux effets chez des enfants musiciens, afin d'estimer comment deux habiletés (la lecture de mots et la dénomination de notes) qui sont automatisées simultanément avec la pratique interfèrent l'une envers l'autre. Les résultats confirment la coexistence des deux effets, le MSE et le RMSE, déjà obtenue avec des musiciens adultes. Mais nos données témoignent surtout de l'importance du concept de la force relative de Cohen, Dunbar, et McClelland (1990). Nous avons montré en effet que la force relative de la lecture de mots et de la dénomination de notes déterminait l'évolution conjointe des deux effets (le MSE et le RMSE), ce qui coïncide avec les suggestions de Grégoire et al. (sous presse) selon lesquelles l'asymétrie généralement observée dans les tâches de type Stroop proviendrait de la force relative des deux traitements en compétition, et pas de l'automaticité de la lecture de mots seulement.

Abstract

In classical Stroop tasks, an asymmetry has been observed with much greater interference of word reading on color naming (Stroop effect) than vice versa (reverse Stroop effect). When a reverse Stroop effect is obtained, it is most of the time at the expense of the Stroop effect itself, as if these two effects could not coexist. By using the musical Stroop paradigm, in which a note picture is presented in a musical staff, with a note name written inside the note picture, Grégoire, Perruchet, and Poulin-Charronnat (in press) have demonstrated that participants exhibited both the musical Stroop effect and the reverse musical Stroop effect. The present study evaluated the joint evolution of the two effects in musician children, in order to assess how two abilities (word reading and note naming) that were concurrently automatized with practice interfered with each other. The results confirmed the co-existence of both the musical Stroop effect and the reverse musical Stroop effect already observed with adult musicians. More importantly, by using a developmental approach, they highlighted the importance of the Cohen, Dunbar, and McClelland (1990)'s concept of relative strength, by showing that the relative strength of word reading and note naming was predictive of the joint evolution of these two effects. The present findings concur with those of Grégoire et al. (in press) to suggest that the classical asymmetry comes from the relative strength of the two competing dimensions involved in Stroop-like paradigms, and not from the automaticity of word reading only.

Introduction

Grégoire, Perruchet, and Poulin-Charronnat (2013, see Akiva-Kabiri & Henik, 2014; Gast, 2014; Moeller & Frings, 2014; Zakay, 2014, for comments; and Grégoire, Perruchet, & Poulin-Charronnat, 2014, for a reply) have demonstrated a new Stroop-like effect, they called the musical Stroop effect (MSE). In the musical Stroop paradigm, the stimuli are composed of a note picture presented at several positions in a musical staff. A name of a note is printed inside the note picture. The written name is either congruent or incongruent with the position of the note in the staff. Adult musicians, who were asked to read the written note names while ignoring the positions of the note in the staff, were slowed down when the written names were incongruent rather than congruent with the note positions. This MSE was interpreted as reflecting the automaticity of note naming in musicians, as Stroop effect is interpreted as reflecting the automaticity of word reading.

It is worth noting that in the musical Stroop paradigm word reading is still present, but its status is inverted: Instead of being the interfering dimension, as usually in classical Stroop-like paradigm, word reading takes the role of the interfered dimension. In classical Stroop terminology, the MSE observed by Grégoire, Perruchet, and Poulin-Charronnat (2013; in press, submitted) is thus a reverse Stroop effect (RSE). The RSE reflects the interference from one competing dimension (e.g., color) on word reading.

The reverse Stroop effect

Although Stroop effect has been widely studied, far few studies have investigated the RSE. Stroop himself (1935) has been the first to investigate the RSE. Contrary to the large interference generated by incongruent words on color naming, Stroop observed only little interference of incongruent ink colors on word reading. Stroop trained participants to name the ink color of incongruent words. He reported an RSE only after extensive practice, but this effect disappeared few days after participants strop practicing. An RSE was obtained only when word reading was strongly degraded (Dunbar & MacLeod, 1984), or when verbal responding was replaced by motor responding (e.g., Blais & Besner, 2006, 2007; Durgin, 2003; Melara & Mounts, 1993). However, other studies using another competing dimension than color, revealed an RSE (e.g., Akiva-Kabiri & Henik, 2012; Grégoire et al., 2013; Palef & Olson, 1975). In Akiva-Kabiri and Henik (2012), musician participants were asked to read the name of notes while hearing a tone that either corresponded to the note or not, or to name auditory tones while ignoring the written note name. While absolute pitch possessors showed a significant congruity effect in the word-reading task, indicating that pitch identification in absolute pitch possessors was impossible to suppress, they were unaffected by the written note name when asked to name the auditory tone.

Palef and Olson (1975) and Akiva-Kabiri and Henik (2012) evaluated conjointly Stroop effect and RSE. Their results seem to indicate that the presence of one of the effect, excluded the presence of the other effect: If Stroop effect is present, RSE is absent and vice versa. This finding was in agreement with the prediction of the Cohen, Dunbar, and MacClelland (1990)'s connectionist model of Stroop effect. However, by using the musical Stroop paradigm, Grégoire et al. (in press) found both effects conjointly, an MSE and what they called a reverse MSE (RMSE), which corresponds to the interference of word on note naming. They observed that the MSE and the RMSE were negatively and positively correlated, respectively, to the relative strength of both abilities (word reading and note naming) measured as the difference between times needed to name notes versus to read words. They run additional analyses in which they classified participants according to the relative strength of the two abilities (RTs in note naming – RTs in word reading). It is worth stressing that for all participants, RTs in note naming was longer than RTs in word reading. The participants were then distributed in three groups: 1) a group called "balanced" for which the difference in strength for both abilities was minimal; 2) a group called "unbalanced", for which the same difference was maximal; and 3) a group called "intermediate", for which the difference was in between. Grégoire et al. (in press) observed that in the balanced group, both the MSE and the RMSE were present and that there was no significant difference in amplitudes between both effects. On the contrary, for the unbalanced group, only the RMSE was significant and virtually no MSE was observed. It appears that more practice makes note naming more interfering, while becoming less susceptible to interference.

A more direct way to evaluate the influence of the relative strength of competing abilities is to vary the level of practice of one of the ability. For instance, in the classical color-word Stroop task, increasing reading skills increases the unbalanced between word and color dimensions. A greater unbalanced should reflect a larger Stroop effect and a reduced RSE.

Evolution of Stroop-like interference with practice

Comalli, Wapner, and Werner (1962) were the first to conduct a life-span study on the evolution of Stroop interference. They tested participants from seven- to 80-year olds, which were classified by age in 11 groups (7, 8, 9, 10, 11, 12, 13, 17-19, 25-34, 35-44, and 65-80 years old). Stroop interference was measured as the difference between times needed to name the color of incongruent color words (e.g., the word BLUE printed in *green* ink) versus color patches. Greatest Stroop interference was found for seven-year-old children, then it decreased up to 17-19 years, remained constant up to 44 years, and increased again after 65 years. Following this study, various data reinforced these findings mainly with the color-word Stroop task (Armengol, 2002; Dash and Dash, 1982; Peru, Faccioli, & Tassinar, 2006; Rand, Wapner, Werner, & McFarland, 1963; Schadler & Thissen, 1981; Schiller, 1966) but also with variants of the Stroop task, such as the picture-word Stroop-like task (Rosinski, Golinkoff, & Kukish, 1975).

The literature reviewed reveals a rather counterintuitive phenomenon: the falling component of the curve. If Stroop interference reflects the automaticity of reading, and if the level of automaticity is a growing function of the amount of practice, then as reading skills improve, the relative strength between reading and a less practiced activity should have become more and more unbalanced in favor of reading, resulting in greater Stroop interference with practice.

However, in all the studies described above, the level of reading practice is entirely confounded with age, and it is possible that at least a part of the observed changes in Stroop interference is due to general factors evolving with age. Comalli et al. (1962) suggested that the amount of interference could be a positive function of the amount of practice, but that this effect would be overshadow by the age-related variations in the ability to inhibit nonrelevant information.

By using the musical Stroop paradigm presented above, Grégoire et al. (submitted) decoupled practice from chronological age, what is simply impossible with word reading due to obvious practical and ethical constraints. They tested children musician with varying levels of musical training, but with similar age, and demonstrated the occurrence of an MSE between the second and third years of musical training. More importantly, they observed a positive relation between the MSE and musical training: The MSE increased with years of musical training. When age differences between groups of different level of practice are reduced, only an increase in Stroop-like interference was observed with increasing practice.

Increasing note-naming skills reduced the unbalanced of the relative strength between word reading and note naming abilities, leading to an increasing MSE.

The present study

To date, in the studies investigating the evolution of Stroop-like effect, the practice was manipulated for only one of the two competing. MacLeod and Dunbar (1988), for instance, used a Stroop-like task in which arbitrary associations were created across repeated experimental sessions, to ensure a better control on practice. Participants were trained to associate a color name to a shape during 20 days. In the experiment of concern here, participants were tested on different tasks on day 1, day 5, and day 20. The authors observed virtually no interference of shape on color naming on Day 1, while the interference became significant on day 5 and continued to increase on Day 20. Shape-naming training reduced the unbalanced in relative strength between color naming and shape naming, leading to an increase interference of shape on color naming.

In a developmental design, Peru, Faccioli, and Tassinari (2006) used the classical color-word Stroop paradigm and tested children from 3 to 10 years old. With increasing age, only word reading was practiced, not color naming. Children of all age were required to name the color, whereas word reading was required only from 6 years upwards. Concerning color naming and as already observed previously, their results showed an increase of Stroop interference up to 7 years and then a continuous decrease. In the word-reading task, all but the 6-year olds demonstrated a significant RSE. Its amplitude was much smaller than the amplitude of the Stroop effect, and the RSE decreased with age.

The present study aimed at assessing how two abilities, becoming concurrently automatized with practice, interfered with each other. By exploiting the musical Stroop paradigm, we evaluated conjointly the MSE and the reverse MSE (RMSE) in musician children with both increasing reading skills and note naming skills (as evaluated with word-reading and note-naming-ability tests). As in Grégoire et al. (in press), two tasks were used. In a word-reading task, the musician children were required to read written note names, while ignoring the position of the note at which they were printed. In a note-naming task, the same musician children were asked to name notes, while ignoring the written note name printed inside. As in Grégoire et al. (in press), it was expected that both the MSE and the RMSE co-occured, and that the relative strength of word reading and note naming was predictive of the pattern of interference.

Method

Participants. One hundred children from the *Conservatoire à Rayonnement Régional* of Dijon took part in the experiment. The children were divided into five groups according to their number of years of musical training (from 1 to 5); one year, $N = 20$, Mean age = 8.20 years, $SD = 0.70$, 10 females; two years, $N = 20$, Mean age = 8.99 years, $SD = 0.76$, 9 females; three years, $N = 21$, Mean age = 10.94 years, $SD = 1.81$, 14 females; four years, $N = 20$, Mean age = 11.01 years, $SD = 0.67$, 13 females; and five years, $N = 19$, Mean age = 12.16 years, $S = 1.38$, 13 females. The children's vision was normal or corrected to normal and written parental consent was obtained for each child.

Materials. The experimental material was composed stimuli consisted of a treble staff with a note picture, which could appear on each of the seven possible positions going from C4 to B4. The name of a note was written inside the note picture. For the congruent condition, the note name was congruent with the note position on the staff (Figure 1a), whereas in the incongruent condition, note name and position were incongruent, with the name written inside the note picture being one of the six other possible note names (e.g., when the note was DO, the written name was LA, SI, RE, MI, FA, or SOL; Figure 1b).

The stimuli used in additional tests of reading and note-naming abilities were designed to be closer from those involved in everyday situations. They are represented in Figure 1c and Figure 1d, respectively.

To prevent the iconic memory of the staff to influence the processing of the following note, the stimuli were randomly displayed at one of four possible positions without immediate repetition at the same location. The four positions were defined as the center of (invisible) rectangles resulting from the exhaustive partitioning of the screen into four quadrants of equal size. Stimuli were printed in black over a white background on a computer screen. Note names appeared in standard uppercase font 14. The treble staff was 7.7 cm wide by 5.1 cm high.

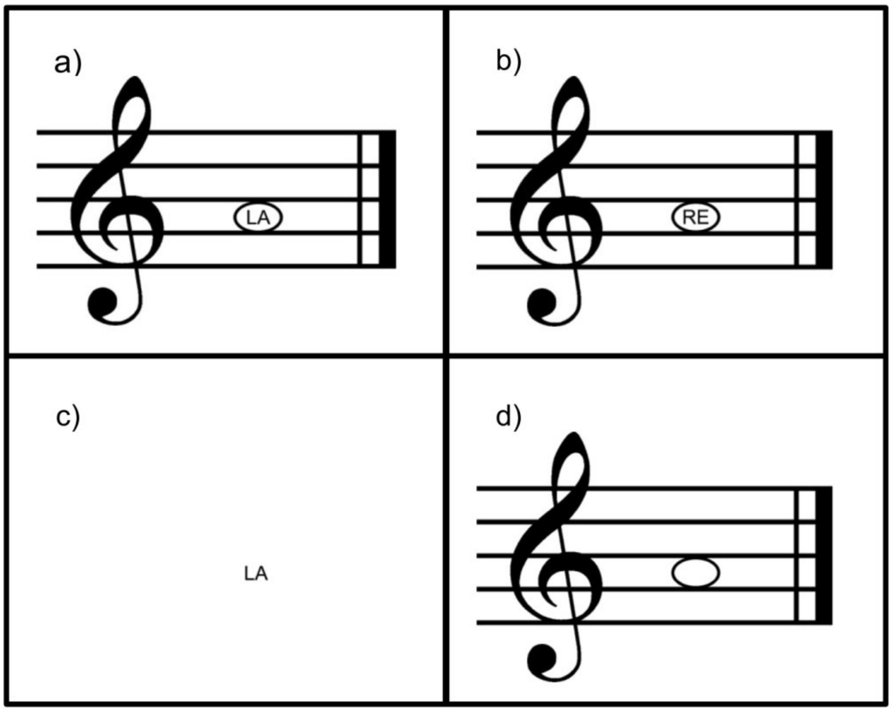

Figure 1. Examples of the different conditions used in the experiment: a) congruent condition; b) incongruent condition; c) reading-ability test; and d) note-naming ability test. Note that in the musical French notation (and several other countries such as Italy and Spain), note names are DO, RE, MI, FA, SOL, LA, SI, instead of the first letters of the alphabet.

Procedure. The musician children had to perform a word-reading task and a note-naming task in succession. The order of presentation of the two tasks was counterbalanced across children. In the word-reading task, the children had to read aloud the printed word while ignoring the note picture. In the note-naming task, the children were asked to name the note while ignoring the word written inside.

For each task, there were two mixed conditions: congruent (Figure 1a) and incongruent (Figure 1b). For each condition, the stimuli appeared six times on each of the 7 locations, resulting in 42 trials per condition, and 84 trials (42 × 2) for each task. On each trial, a fixation cross was displayed for 1 s at the center of the screen before the apparition of the stimulus, which stayed on the screen until participant's response. The interval between the response and the next trial was 1s. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of note locations or note names. They were displayed as four blocks of 21 trials each with a self-paced break between blocks.

The experimental session was immediately followed by two additional tests, which were run in counterbalanced order. One test was a reading-ability test, in which the children had to read note names (Figure 1c). The other test was a note-naming ability test, in which the children had to name notes (Figure 1d). Each test included 21 trials. The trials were pseudo-randomly ordered for each participant, excluding immediate repetitions of words or notes.

Whatever the tasks, participants were encouraged to respond as fast and as accurately as possible throughout the session. The response times (RTs) were recorded by a voice key. During the session, the experimenter noted error responses and voice-key dysfunctions. After the experiment, the children filled out a questionnaire about their musical training.

Results

Ability tests

Voice-key dysfunctions led to exclude 3.83% of the data. There was virtually no error in the word reading-ability test, $M = 0.24$ %, $SD = 1.25$. The proportion of errors in the note-naming ability test was higher than the word reading-ability test, $M = 2.46\%$, $SD = 5.03$ (Table 1). An ANOVA on children's errors performed, with Task (word reading, note-naming) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, showed a significant main effect of task ,with more errors in the note-naming task compared with the word-reading task, $F(1, 95) = 18.83$, $p < .001$, $\eta_p^2 = .165$. There was no main effect of years of musical training, $F(4, 95) = 1.83$, $p = .13$ and no Task × Years of Musical Training interaction, $F(4, 95) = 1.81$, $p = .133$.

RTs for correct responses beyond three standard deviations of the mean (0.85%) were removed. The remaining data are shown in Figure 2, top panel. An ANOVA on children's correct RTs performed, with Task (word reading, note-naming) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, showed a significant main effect of task, with shorter RTs for the word-reading task compared with the note-naming task, $F(1, 95) = 99.64$, $p < .001$, $\eta_p^2 = .512$, and a significant main effect of years of musical training, with faster RTs as the number of years of musical training increased, $F(4, 95) = 28.40$, $p < .001$, $\eta_p^2 = .545$.

These two main effects were qualified by a significant Task × Years of Musical Training interaction, $F(4, 95) = 20.32$, $p < .001$, $\eta_p^2 = .461$. For the word-reading ability test (Figure 2, top panel, left), there was a significant main effect of years of musical training, $F(4, 95) = 10.49$, $p < .001$, $\eta_p^2 = .306$, with a significant decreasing linear trend, $p < .001$. The pattern of results was observed for the note-naming ability test (Figure 2, top panel, right), with a significant main effect of years of musical training, $F(4, 95) = 25.56$, $p < .001$, $\eta_p^2 = .518$, and a significant decreasing linear trend, $p < .001$.

Note that there was no evidence for a speed-accuracy trade-off in the note-naming ability test, as shown by the significant positive correlation between errors and RTs, $r(98) = .263$, $p = .008$.

Table 1
Proportions of errors as a function of years of musical training and condition

	Years of musical training				
	1 year	2 years	3 years	4 years	5 years
Word reading ability test	0.25 (1.12)	0.71 (2.33)	0 (0)	0.24 (1.17)	0 (0)
Note-naming ability test	5.14 (8.68)	1.74 (3.34)	1.60 (2.76)	1.73 (3.29)	2.11 (4.19)
Word-reading congruent	0.12 (0.55)	0.12 (0.53)	0.47 (1.26)	0.24 (0.74)	0.13 (0.57)
Word-reading incongruent	0.47 (1.23)	0.77 (2.37)	0.71 (1.38)	1.35 (2.46)	1.48 (2.43)
Note-naming congruent	3.14 (4.46)	1.97 (3.74)	1.60 (4.15)	0.24 (075)	1.06 (3.52)
Note-naming incongruent	9.23 (9.11)	10.12 (6.98)	7.55 (6.54)	6.11 (5.09)	7.99 (7.86)

Word reading (MSE)

Voice-key dysfunctions led to exclude 3.46% of the data. A small proportion of errors was observed. An ANOVA on children's reading errors performed with Congruity (congruent, incongruent) as a within-subject variable showed a significant main effect of congruity, $F(1, 95) = 13.52$, $p < .001$, $\eta_p^2 = .125$, with more errors in the incongruent condition than in the congruent condition, attesting to an MSE. There was no main effect of years of musical training, $F(4, 95) = 0.70$, $p = .592$. Although the MSE for errors globally increased with the number of years of musical training, no significant Congruity × Years of musical training interaction was observed, $F(4, 95) = 1.13$, $p = .347$, and the linear trend failed to reach significance, $p = .091$.

RTs for correct responses beyond three standard deviations of the mean (1.00%) were removed. The remaining data are shown in Figure 2, bottom panel, left. An ANOVA on RTs was carried out with Congruity (congruent, incongruent) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable. There was a significant main effect of congruity, $F(1, 95) = 60.06$, $p < .001$, $\eta_p^2 = .387$,with longer RTs in the incongruent than in the congruent condition, and a significant main effect of years of musical training, $F(4, 95) = 8.68$, $p < .001$, $\eta_p^2 = .268$, with decreasing RTs with years of musical training.

Although the MSE dropped from 18.16 ms to 30.67 ms from one to five years of musical training, no Congruity × Years of Musical Training interaction, $F(4, 95) = 0.43$, $p = .786$, and no significant linear trend, $p = .238$, were observed. However, because the pattern of RTs across congruity conditions for each group was the principal concern of the present study, further *t*-test analyses were performed. All the groups showed significantly longer RTs in the incongruent condition than in the congruent condition, $t(19) = 2.536$, $p = .020$, $d = 0.17$, $t(19) = 3.398$, $p = .003$, $d = 0.41$, $t(20) = 5.231$, $p < .001$, $d = 0.32$, $t(19) = 2.877$, $p = .010$, $d = 0.52$, and $t(18) = 4.716$, $p < .001$, $d = 0.43$, for one, two, three, four, and five years of musical training, respectively.

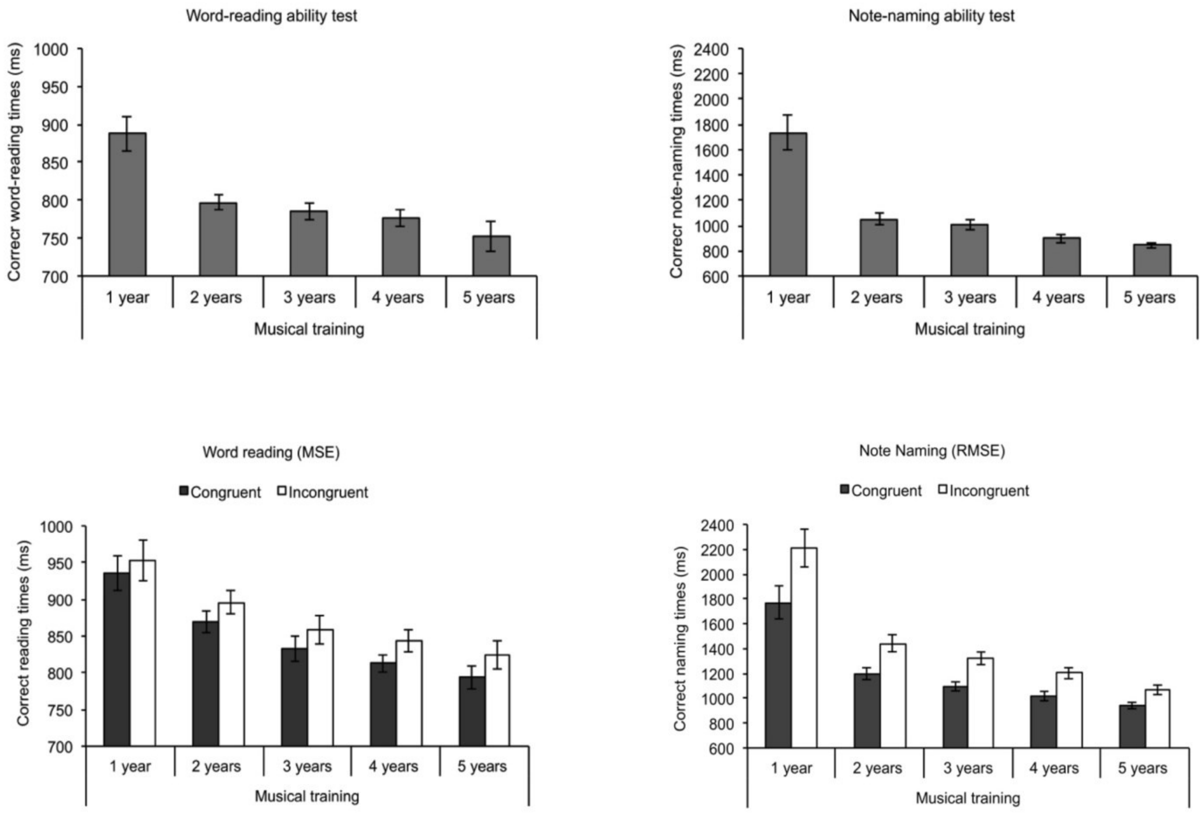

Figure 2. Correct response times for the word-reading ability test (top panel, left) and the note-naming ability test (top panel, right) as a function of Years of Musical Training; correct response times for the word-reading task (bottom panel left) and for the note-naming task (bottom panel, right) as a function of Congruity and Years of Musical Training. Error bars indicate standard errors.

Note naming (RMSE)

Voice-key dysfunctions led to exclude 5.44% of the data. An ANOVA on errors, carried out with Congruity (congruent, incongruent) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, showed a significant main effect of Congruity, $F(1, 95) = 89.02$, $p < .001$, $\eta_p^2 = .484$, with more errors in the incongruent condition than in the congruent condition, attesting to an RMSE, and no main effect of years of musical training, $F(4, 95) = 1.52$, $p = .202$. Neither the Congruity × Years of Musical Training interaction was not significant, $F(4, 95) = 0.38$, $p = .82$, nor the linear trend, $p = 905$.

Correct RTs beyond three standard deviations of the mean (1.46%) were removed. The remaining data are shown in Figure 2, bottom panel, right. As for the errors, an ANOVA on RTs was carried out with Congruity (congruent, incongruent) as a within-subject variable and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable. There was a significant main effect of congruity, $F(1, 95) = 295.68$, $p < .001$, $\eta_p^2 = .757$, with longer RTs for the incongruent than for the congruent condition, and a significant main effect of years of musical training, $F(4, 95) = 26.04$, $p < .001$, $\eta_p^2 = .523$, with shorter RTs as the number of years of musical training increased. These two main effects were qualified by a significant Congruity × Years of Musical Training interaction, $F\ (4, 95) = 13.85$, $p < .001$, $\eta_p^2 = .368$, with a significant linear trend, p < .001, indicating that the difference between incongruent and congruent (RMSE) linearly decreased from one to five years of musical training.

Further *t*-test analyses showed that all the groups showed significantly longer RTs for the incongruent condition than for the congruent condition, $t(19) = 7.942$, $p < .001$, $d = 0.70$, $t(19) = 8.731$, $p < .001$, $d = 0.97$, $t(20) = 10.82$, $p < .001$, $d = 1.19$, $t(19) = 8.579$, $p < .001$, $d = 1.02$, and $t(18) = 7.808$, $p < .001$, $d = 0.91$, for one, two, three, four, and five years of musical training, respectively.

Comparing MSE and RMSE

An ANOVA on errors, carried out with Task (word reading, note naming) and Congruity (congruent, incongruent) as a within-subject variables and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, revealed a significant main effect of task, $F(1, 95) = 91.30$, $p < .001$, $\eta_p^2 = .49$, with more errors in the note-naming task than in the word-reading task, and a significant main effect of congruity, $F(1, 95) = 98.67$, $p < .001$, $\eta_p^2 = .509$, with more errors in the incongruent condition than in the congruent condition. No main effect of years of musical training was observed, $F(4, 95) = 0.96$, $p = .433$. The Task × Congruity interaction was significant, $F(1, 95) = 66.83$, $p < .001$, $\eta_p^2 = .413$, indicating that the RMSE is larger than the MSE. There was no other significant effect (Task × Years of Musical Training interaction, $F(4, 95) = 2.05$, $p = .094$; Congruity × Years of Musical Training interaction, $F(4, 95) = 0.49$, $p = .745$; Task × Congruity × Years of Musical Training interaction, $F(4, 95) = 0.39$; $p = .814$).

An ANOVA on RTs, carried out with Task (word reading, note naming) and Congruity (congruent, incongruent) as a within-subject variables and Years of Musical Training (1, 2, 3, 4, 5) as a between-subject variable, revealed a significant main effect of task, $F(1, 95) = 215.09$, $p < .001$, $\eta_p^2 = .694$, with shorter RTs in the word-reading task compared with the note-naming task, a significant main effect of congruity, $F(1, 95) = 366.24$, $p < .001$, $\eta_p^2 = .794$, with longer RTs in the incongruent condition compared with the congruent condition, and a significant main effect of years of musical training, $F(4, 95) = 27.01$, $p < .001$, $\eta_p^2 = .532$, with RTs decreasing with the number of years of musical training. The three main effects were qualified by several interactions. There was a significant Task × Congruity interaction, $F(1, 95) = 209.75$, $p < .001$, $\eta_p^2 = .688$, a significant Task × Years of Musical Training interaction, $F(4, 95) = 22.40$, $p < .001$, $\eta_p^2 = .485$, a significant Congruity × Years of Musical Training interaction, $F(4, 95) = 12.87$, $p < .001$, $\eta_p^2 = .351$, and finally a significant Task × Congruity × Years of Musical Training interaction, $F(4, 95) = 13.37$, $p < .001$, $\eta_p^2 = .360$. This three-way interaction indicated that the MSE did not increase with years of musical training, whereas the RMSE showed a significant linear decrease.

Individual differences

As in Grégoire et al. (in press) the relative strength (see Cohen et al., 1990) of both word-reading and note-naming abilities was computed to evaluate its influence on the musician children performance. The relative strength of word-reading and note-naming abilities for a given musician child can be given by the difference in RTs between the two ability tests. A participant P1 whose the difference between note naming and word reading is larger than for a participant P2 can be taken as having a stronger imbalance between note-naming and word-reading abilities than participant P2. Is this difference actually predictive of the pattern of interference?

The response is clearly positive. There was a negative correlation between the relative strength of the two abilities and the MSE, $r(98) = -.264$, $p = .008$, which remained significant when age was partialled out, $r(97) = -.274$, $p = .006$, and there was a positive correlation between the relative strength of the two abilities and the RMSE, $r(98) = .620$, $p < .001$, which remained significant when age was partialled out, $r(97) = .557$, $p < .001$.

To provide a more complete picture of the resulting pattern, participants were divided into five groups of 20 participants each, along the relative strength dimension. At one extreme of the resulting classification, the mean difference in speed between the two abilities was moderate. Participants exhibiting this pattern will be called *balanced+*. At the other extreme, the difference in strength between the two abilities is maximal. Participants exhibiting this pattern will be called *balanced-*. The remaining participants are in-between.

Figure 3 shows the pattern of interference for each group. Overall, an ANOVA performed on the amount of interference with Task (word reading, note naming) as a within-subject variable and Group (balanced+, 3 levels of intermediate balanced-) as a between-subject variable, revealed a significant main effect of Task, $F(1, 95) = 219.18$, $p < .001$, $\eta_p^2 = .698$, with a larger RMSE than MSE, a significant main effect of Group, $F(4, 95) = 11.48$, $p < .001$, $\eta_p^2 = .326$, with an interference that increased the balanced+ to the balanced- group, and, crucially, a significant Group × Task interaction, $F(4, 95) = 14.62$, $p < .001$, $\eta_p^2 = .381$, with a significant linear trend, $p < .001$, indicating that the difference between MSE and RMSE linearly increased from the balanced+ group to the balanced- group.

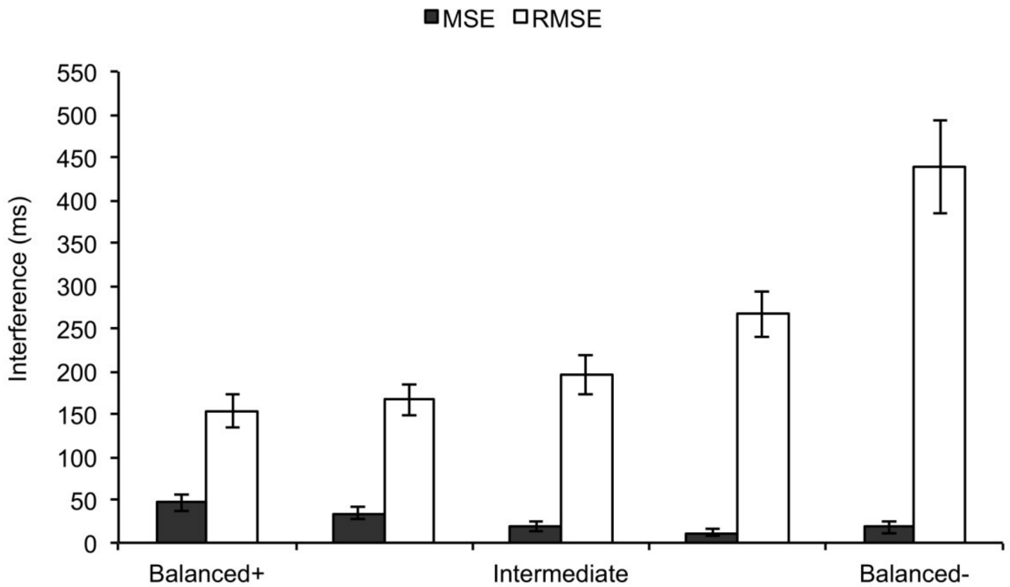

Figure 3. Amount of interference (RTs in incongruent condition minus RTs in congruent condition) in the word-reading and note-naming tasks for five groups of children musicians, sorted according to their performance in independent tests of reading and note-naming ability. Balanced+ group: note naming was slightly longer than word reading; Balanced- group: note naming was largely slower than word reading; Intermediate group: in-between. Error bars indicate standard errors.

Discussion

The present study used the musical Stroop paradigm to evaluate the concurrent evolution of MSE and RMSE. Musician children performed two tasks: A word-reading task, in which they had to read note names printed inside a note picture while ignoring the position of the note in the staff and a note-naming task, in which they had to name notes while ignoring the note names written inside. In agreement with Grégoire et al. (in press), the results showed that both effects, MSE and RMSE, were present together and that the relative strength of both abilities (word reading and note naming) was predictive of the pattern of interference observed in musician children's performance.

Relation between Stroop-like interference and practice

Although the MSE increased from the first to the fifth years of musical training, this positive relation was not significant. In the present experiment, the five groups of participants showed a significant MSE. This was not the case in Grégoire et al. (submitted), in which the MSE appeared between the second and third year. However, the children with one year of musical training tested in the present study were younger than in Grégoire et al., making possible that they possessed lower reading skills. Reading in these children might be more susceptible to interference because less automatized. In addition, this difference might also be related to differences regarding the sampling of participants. In the present study, musician children came from the music conservatory of Dijon, the objective of which is to train professional musicians. In Grégoire et al. (Experiment 2), musician children were recruited from a larger number of musical schools from the suburban area of Dijon, making the musical levels more heterogeneous and certainly of a lower quality. This difference and its consequence is reminiscent of the study of Armengol (2002), who also observed a one-year difference in the appearance of the classical Stroop effect between private and public schools.

A second important finding is the negative relation observed between RMSE and musical training. When the musician children were asked to name the notes, a strong RMSE was observed, however as years of musical training increased, the RMSE significantly decreased. In the present study, both note-naming skills and reading skills were concurrently manipulated. The consequence of varying reading skills is that the musician children participated into the experiment have different ages, children with more reading skills being older. The decrease observed in RMSE could be perceived as a replication of the decrease observed in previous developmental studies (e.g., Comalli et al., 1962) and could be interpreted as reflecting age changes. However, if that were the case, a similar decrease should have been observed with the MSE. The increase in MSE, although not significant, rules out this interpretation.

Another interpretation is that with increasing practice, note naming becomes more and more irrepressible but also less susceptible to interference. There is a long-standing debate regarding the question of whether the different properties commonly attributed to automatisms evolve in parallel (e.g., Johnston & Dark, 1986; Paap & Odgen, 1981). Regarding the speed of processing and the independence from attentional resources, there is a consensus that more practice leads to a gradual and continuous improvement. Earlier results on the evolution of the Stroop effect could have suggested that irrepressibility have to be dissociated from other properties. Our results suggest that such a conclusion could be unwarranted, and presumably, irrepressibility, as other characteristics of automatisms such as the susceptibility to interference, is a linear function of practice.

Importance of the relative strength of two abilities

As already discussed by Grégoire et al. (in press), the co-occurrence of both the MSE and the RMSE cannot be accounted for by the Cohen et al.'s (1990) connectionist model. Cohen et al.'s claimed that "the relative strength, compared with a competing pathway [that is] important in determining whether a process will produce or be subject to interference in a Stroop-like task" (p. 348). Grégoire et al. (in press) proposed a softened version of this claim: The relative strength, compared with a competing pathway would be important in determining the relative propensity of the stronger process to interfere on, and be interfered by another, weaker process; the final pattern taking the form of a trade-off between the amount of interference affecting each pathway. The present study gives further evidence concerning the importance of the relative strength of both the interfered and the interfering dimensions in a Stroop-like task.

As in Grégroire et al. (in revision), the strength of word reading, assessed by the speed obtained in the ability test, was much stronger than the strength of note naming (Figure 2). As a consequence, word reading elicited much greater interference on note naming then the reverse. However, individual differences indicated that the relative strength of word reading and note naming, calculated by subtracting reading times from naming times in the ability tests, was correlated negatively with the MSE, and positively with the RMSE. When the musician children were distributed in five groups according to relative strength, a significant linear trend indicated that the difference between MSE and RMSE linearly increased from the balanced group to the unbalanced group. This finding reinforces the conclusion of Grégoire et al. (in press), according with the relative strength of two competing abilities determined the pattern of interference.

In previous color-word Stroop studies, investigating the RSE, it has been shown to be very difficult to observe an RSE. In light of our results, the main reason for this is not so much related to the automaticity of word reading but more to the “non-automaticity” of the competing dimension. As soon as two competing dimensions are practiced dimension with a certain level of automaticity, then both Stroop effect and RSE can be concurrently observed and there is no longer what was called *Stroop asymmetry*. Because the focus was put on word reading as the interfering dimension, this asymmetry is very likely due to the usual unbalanced of relative strength between reading and the other competing dimension. However, in agreement with our previous data, the present results demonstrated that even reading irrepressibility might be susceptible to interference.

References

Akiva-Kabiri, L. & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology, 59*(5), 272-278.

Akiva-Kabiri, L. & Henik, A. (2014). Additional insights: Commentary on “ The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269-278). *Experimental Psychology, 61*(1), 75-77.

Armengol, C. G. (2002). Stroop test in Spanish: Children's norms. *Clinical Neuropsychologist, 16*(1), 67-80.

Blais, C., & Besner, D. (2006). Reverse stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1345-1353.

Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review, 14*(3), 466-469.

Braet, W., Noppe, N., Wagemans, J., & Op de Beeck, H. (2011). Increased Stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology, 64*(3), 596-607.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332-361.

Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interfernce effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology, 100*, 47-53.

Dash, J., & Dash, S. (1982). Cognitive developmental studies of the Stroop phenomena: Cross-sectional and longitudinal data. *Indian Psychologist, 1*(1), 24-33.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 622-639.

Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics, 65*(3), 367-378.

Ehri, L. C., & Wilce, L. S. (1979). Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology, 27*, 352-364.

Gast, A. (2014). What is learned, and when? Commentary on “ The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and

B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269-278). *Experimental Psychology, 61*(1), 71-74.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*(4), 269-278.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014). Is the musical Stroop effect able to keep its promises? A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology, 61*(1), 80-83.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (in press). About the unidirectionality of interference: Insight from the musical Stroop effect. *Quarterly Journal of Experimental Psychology.*

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (submitted). How does Stroop interference vary with practice? A reappraisal from the musical Stroop paradigm.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. Annual Review of Psychology, 37, 43–75.

Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21*(5), 627-645.

Moeller, B., & Frings, C. (2014). How automatic is the musical the musical Stroop effect ? Commentary on " The musical Stroop effect: Opening a new avenue to research on automatisms" by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269-278). *Experimental Psychology, 61*(1), 68-70.

Paap, K. R., & Ogden, W. C. (1981). Letter encoding is an obligatory but capacity-demanding operation. *Journal of Experimental Psychology : Human Perception and Performance, 7*, 518–527.

Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task.*Canadian Journal of Psychology, 29*(3), 201-209.

Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes de Biologie, 144*(1), 45-62.

Rand, G., Wapner, S., Werner, H., & McFarland, J. H. (1963). Age differences in performance on the Stroop color-word test. *Journal of Personality, 31*, 534-558.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46*, 247-253.

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9*(2), 132-141.

Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology, 72*(1), 105-108.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior, 13*, 57-74.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Zakay, D. (2014). Can the ‘‘musical Stroop’’ task replace the classical Stroop task? Commentary on “ The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269-278). *Experimental Psychology, 61*(1), 78-79.

<u>Etude 5</u>

Un automatisme non sollicité résiste-t-il aux ravages du temps?

Introduction

Le comportement automatique, notamment celui qui renvoie aux habiletés, a coutume d'être assimilé à une certaine immutabilité. Il est généralement admis, par exemple, que savoir faire du vélo « ne se perd pas ». Anderson (1983) a d'ailleurs argué que les apprentissages procéduraux ne s'oubliaient jamais. Mais qu'en est-il d'un automatisme cognitif ? Est-ce qu'un traitement comme la lecture s'exécuterait toujours sur un mode automatique après plusieurs années sans pratique ? Il serait utopique de vouloir répondre à cette dernière question étant donné l'ubiquité des mots dans notre environnement quotidien. En revanche, la dénomination de notes, qui est un traitement automatisé chez les musiciens (voir Grégoire, Perruchet, & Poulin-Charronnat, 2013), représente un champ d'investigation plus approprié pour explorer cette problématique. En effet, le contrôle de l'activité musicale (et particulièrement de la dénomination de notes) est beaucoup plus aisé que celui de la lecture, car les notes de musique ne sont pas omniprésentes dans notre environnement. Il serait donc tout à fait possible de déterminer si le traitement automatique de la dénomination de notes persiste après un arrêt prolongé de la pratique en utilisant le paradigme du Stroop musical (Grégoire et al., 2013).

Peu d'études se sont attachées à évaluer le devenir de processus, qui ont été automatisés dans le passé, après un arrêt total et continu de la pratique de l'activité. Kolers (1976) a demandé à des étudiants de lire des textes dont la typographie était inversée un an après les avoir soumis à la même tâche. Il constata que les performances des sujets étaient meilleures lors de la seconde passation, même pour des textes inversés qui n'avaient pas été lus lors de la première passation. Ainsi, l'habileté à décoder des informations graphémiques présentées de manière inhabituelle persiste après plus d'un an sans pratique. MacLeod et Dunbar (1988) ont assigné un nom de couleur spécifique à des formes géométriques non-familières. Ils ont ensuite demandé aux participants d'apprendre ces noms au cours de plusieurs sessions de pratique. Après cinq jours d'entraînement, la forme avait un effet sur la dénomination de la couleur dans une tâche de type Stroop, et cet effet persistait trois mois plus tard malgré une absence totale de pratique de la dénomination de forme (Expérience 2). Les deux études précédentes semblent indiquer que l'automaticité d'un traitement serait susceptible de perdurer même si l'activité n'est pas pratiquée pendant plusieurs mois. Par ailleurs, Hasher et Zacks (1979) ont avancé que la pratique n'apportait aucun bénéfice aux traitements automatiques, ce qui pourrait suggérer qu'une absence de pratique ne leur serait pas préjudiciable.

L'objectif de cette étude est d'observer si des sujets qui ont reçu une formation de solfège soutenue, leur ayant permis d'automatiser la lecture de notes, ont conservé des traces de leur apprentissage après avoir arrêté toute activité musicale depuis un certain nombre d'années. Nous nous intéresserons uniquement à des sujets qui ont cessé toute activité musicale depuis au moins trois ans, ce qui représente un délai beaucoup plus important que celui retenu dans les travaux antérieurs qui se sont penchés sur la problématique de cette étude. Nous avons repris les mêmes conditions expérimentales que dans l'Etude 4 afin d'observer si des effets similaires étaient obtenus. Nous voulions vérifier notamment si les corrélations entre la force relative des deux traitements (lecture de mots et dénomination de notes) et chacun des deux effets de type Stroop (le MSE et le RMSE) étaient répliquées.

Méthode

Participants. Vingt-quatre adultes ont participé volontairement à cette étude en répondant à une annonce publiée dans la presse locale de Dijon (Le Bien Public, Dijon Mag) ou sur des réseaux sociaux (Facebook, Twitter, uB link). Ils étaient tous de langue maternelle française, avaient une vue normale ou corrigée, et n'éprouvaient aucun problème de perception des couleurs. Les participants ont suivi des cours de solfège dans un conservatoire ou une école de musique pendant au moins cinq ans. Ils ont cessé toute pratique musicale (instrument, chant, lecture de partition) depuis au moins trois ans. Deux sujets furent retirés des analyses sur la base de leurs très mauvaises performances dans le test d'habileté en dénomination de notes. Les 22 participants restants (16 femmes; 34.95 ans en moyenne, ET = 12.09) ont une moyenne de 12.68 ans de pratique musicale (ET = 7.74). Leur durée d'arrêt moyenne est de 14.30 ans (ET = 9.12).

Matériel et procédure. Pour la première partie de la passation, le matériel expérimental et la procédure étaient similaires à ceux de l'Etude 4, excepté que tous les sujets commençaient par la tâche de lecture de mots (pour éviter de réactiver des connaissances musicales en commençant par la tâche de dénomination de notes). Le test d'habiletés de lecture fut également réalisé avant le test d'habiletés en dénomination de notes.

Les sujets effectuèrent ensuite les deux tâches du paradigme de Stroop couleur-mot. Nous voulions ainsi évaluer les effets du Stroop classique (avec la tâche de dénomination de couleur) et du Stroop inversé (avec la tâche de lecture de mots) afin d'observer s'ils étaient corrélés avec le MSE et le RMSE. Nous ne présenterons pas les résultats obtenus avec ce second paradigme car ils relèvent d'une problématique qui ne s'inscrit pas dans le projet de thèse.

Résultats

Les tests d'habiletés en dénomination de notes

Deux participants furent retirés des analyses sur la base de leurs mauvaises performances : un sujet a obtenu un temps de réponse moyen supérieur à trois écarts-types de la moyenne, et l'autre n'avait aucune réponse correcte. Nous présenterons ainsi les résultats préliminaires obtenus avec 22 participants.

Les dysfonctionnements de la clé vocale ont conduit à exclure 3.57% des données. Aucune erreur ne fut observée dans le test d'habileté de lecture. La proportion d'erreurs dans le test d'habiletés en dénomination de notes était plus élevée, $M = 1.32\%$, $SD = 3.03$, mais n'atteignait pas la significativité, $t(21) = 2.04$, $p = .054$, $d = .645$. Les TRs des réponses correctes au-delà de trois écarts-types de la moyenne (0.45%) furent retirés. Les TRs corrects étaient significativement plus courts pour la lecture de mots que pour la dénomination de notes, $M = 710.63$, $SD = 59.85$ vs. $M = 894.41$, $SD = 165.28$, respectivement, $t(21) = 6.66$, $p < .001$, $d = 1.55$. Nous n'avons pas constaté de *trade-off* entre la vitesse et la précision dans la tâche de dénomination de notes, comme le montre la corrélation positive entre la proportion d'erreurs et les TRs, $r(20) = .734$, $p < .001$. Ce résultat s'explique par le fait que les seuls sujets qui ont commis des erreurs dans cette situation (4 participants sur 22) étaient également les sujets les plus lents.

Les tâches du paradigme de Stroop musical

Les dysfonctionnements de la clé vocale ont amené à enlever 4.14% des données. Peu d'erreurs furent observées dans la tâche de lecture de mots (0.23%), mais une proportion plus importante fut obtenue dans la tâche de dénomination de notes (2.38%). Une ANOVA sur les erreurs de lecture avec les facteurs Congruité (congruente, incongruente) et Tâche (lecture de mots, dénomination de notes) comme variables intra-sujets révélait des effets principaux significatifs de la Congruité, $F(1,23) = 15.67$, $p < .001$, $\eta_p^2 = .405$, et de la Tâche, $F(1, 23) = 13.83$, $p = .001$, $\eta_p^2 = .376$, de même qu'une interaction Congruité × Tâche significative, $F(1,23) = 11.28$, $p = .003$, $\eta_p^2 = .329$. Dans la tâche de dénomination de notes, les sujets faisaient significativement plus d'erreurs dans la condition incongruente (4.18%) que dans la condition congruente (0.68%), $t(21) = 3.80$, $p < .001$, $d = 1.12$. En revanche, dans la tâche de lecture de mots, bien que le pourcentage d'erreurs étaient plus important dans la condition incongruente (0.33%) que dans la condition congruente (0.11%), la différence n'atteignait pas la significativité, $t(21) = 1.41$, $p = .174$. L'effet de la congruité (évalué par la différence entre la condition incongruente et la condition congruente) était significativement plus fort dans la tâche de dénomination de notes (3.50%) que dans la tâche de lecture de mots (0.21%), $t(21) = 3.43$, $p = .003$, $d = 1.11$.

Les TRs des réponses correctes au-delà de trois écarts-types de la moyenne (1.09%) furent retirés. Les données restantes sont représentées sur la Figure 1. Une ANOVA sur les TRs corrects avec les facteurs Congruité (congruente, incongruente) et Tâche (lecture de mots, dénomination de notes) comme variables intra-sujets révélait des effets principaux significatifs de la Congruité, $F(1,23) = 83.40$, $p < .001$, $\eta_p^2 = .784$, et de la Tâche, $F(1,23) = 50.73$, $p < .001$, $\eta_p^2 = .688$, ainsi qu'une interaction Congruité × Tâche significative, $F(1,23) = 54.10$, $p < .001$, $\eta_p^2 = .701$. Les TRs dans la condition incongruente étaient significativement plus importants que dans la condition congruente, aussi bien dans la tâche de lecture de mots, $t(21) = 3.66$, $p = .001$, $d = .148$, que dans la tâche de dénomination de notes, $t(21) = 9.74$, $p < .001$, $d = .728$, attestant de la présence d'un MSE et d'un RMSE, respectivement. L'amplitude du MSE pouvait sembler relativement faible (11.88 ms), mais l'effet était présent chez 19 des 22 participants (descriptivement, tous les sujets montraient le

RMSE). Par ailleurs, le RMSE était considérablement plus fort que le MSE (165.36 ms vs. 11.88 ms), $t(21) = 8.28$, $p < .001$, $d = 2.81$.

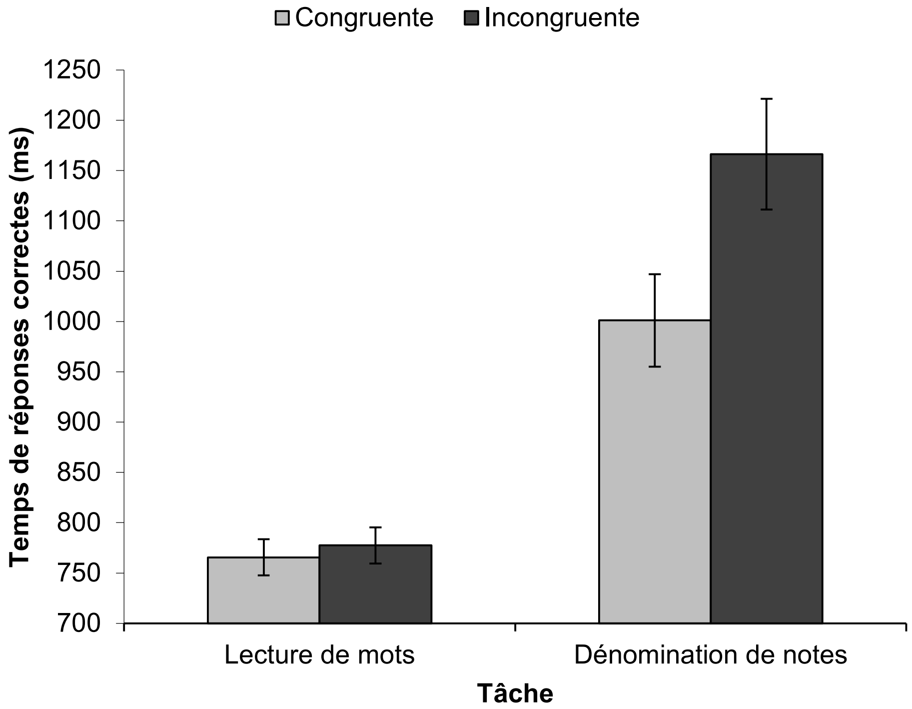

Figure 1. Temps de réponses correctes en fonction de la Tâche et de la Congruité. Les barres d'erreurs indiquent les erreurs standards.

Différences individuelles

Nous avons constaté une absence de corrélation entre le MSE et la durée d'arrêt de la pratique musicale, $r(20) = .052$, $p = .820$, ce qui pourrait indiquer que l'arrêt de l'activité musicale n'a pas d'effet sur l'automaticité de la dénomination de notes. Néanmoins, la taille de notre échantillon ($N = 22$) n'est pas suffisante pour apporter des conclusions convaincantes. De plus, l'interprétation de ce résultat semble difficile dans la mesure où nous n'avons pas d'informations sur le niveau d'automaticité (de la dénomination de notes) des sujets au moment où ils ont cessé de pratiquer la musique.

Une corrélation positive fut observée entre le MSE et la durée de pratique musicale mais elle n'est pas significative, $r(20) = .153$, $p = .497$. Nous avons également obtenu une corrélation négative tendancielle entre le MSE et le RMSE, $r(20) = -.410$, $p = .058$.

Nous avons réitéré les analyses effectuées dans l'Etude 2 pour vérifier si le pattern d'interférence observé dépendait de la force relative des deux traitements en compétition, comme le suggère le modèle de Cohen, Dunbar, et McClelland (1990). Nous avons estimé la force des traitements liés à la dénomination de notes et à la lecture de mots à partir des TRs mesurés dans les tests d'habiletés. La force relative de la dénomination de notes et de la lecture de mots pour un sujet donné est obtenue en calculant la différence de TRs dans les deux tâches. Les TRs en dénomination de notes sont plus lents qu'en lecture pour tous les participants, mais les différences sont très variables en fonction des sujets (de 9.57 ms à 425.76 ms). Comme dans l'Etude 2, nous avons constaté une corrélation négative entre la force relative des deux traitements et la quantité d'interférence dans la tâche de lecture de mots, $r(20) = -.674$, $p < .001$, et une corrélation positive entre la force relative des deux traitements et la quantité d'interférence dans la tâche de dénomination de notes, $r(20) = .512$, $p = .015$. Afin d'apporter une représentation plus complète de ces deux corrélations, nous avons constitué trois groupes de participants (de taille approximativement égale) selon la force relative des deux traitements. A une extrémité de cette classification, la différence de vitesse entre les deux tests d'habiletés est modérée (de 9.57 ms à 92.26 ms). Ce groupe de sujets ($N = 8$) est appelé *équilibré*. A l'autre extrémité, la différence de vitesse entre les deux tests est maximale (de 231.07 ms à 425.76 ms). Ce groupe de participants ($N = 7$) est qualifié de *non-équilibré*. Les sujets restants (*intermédiaire*, $N = 7$) se situaient entre les deux groupes précédents.

La Figure 2 montre le pattern d'interférence pour chaque groupe. Une ANOVA sur la quantité d'interférence avec le facteur Groupe (équilibré, intermédiaire, non-équilibré) comme variable inter-sujets et le facteur Tâche (lecture de mots, dénomination de notes) comme variable intra-sujets indiquait une absence d'effet principal du Groupe, $F(2, 19) = 1.22$, $p = .318$, un effet significatif de la Tâche, $F(1, 19) = 83.24$, $p < .001$, $\eta^2_p = .814$, et une interaction Groupe × Tâche tendancielle, $F(2, 19) = 2.92$, $p = .079$, $\eta^2_p = .235$. Pour le groupe équilibré, le MSE et le RMSE étaient significativement au-dessus de la chance, $t(7) = 5.69$, $p < .001$, $d = 2.15$, et $t(7) = 6.48$, $p < .001$, $d = 2.45$, respectivement, et le RMSE apparaissait plus important que le MSE, $t(7) = 5.42$, $p < .001$, $d = 2.05$. Des résultats analogues furent observés pour le groupe intermédiaire. En effet, le MSE et le RMSE étaient significativement au-dessus de la chance, $t(6) = 4.76$, $p = .003$, $d = 1.94$, et $t(6) = 5.26$, $p = .002$, $d = 2.15$, respectivement, et le RMSE se révélait plus fort que le MSE, $t(6) = 4.62$, $p = .004$, $d = 1.89$. En revanche, pour le groupe non-équilibré, le MSE ne différait pas significativement de la chance, $t(6) = 0.35$, $p = .793$, contrairement au RMSE, $t(6) = 6.37$, $p < .001$, $d = 2.60$. Le RMSE était toujours substantiellement supérieur au MSE, $t(6) = 5.88$, $p = .001$, $d = 2.40$.

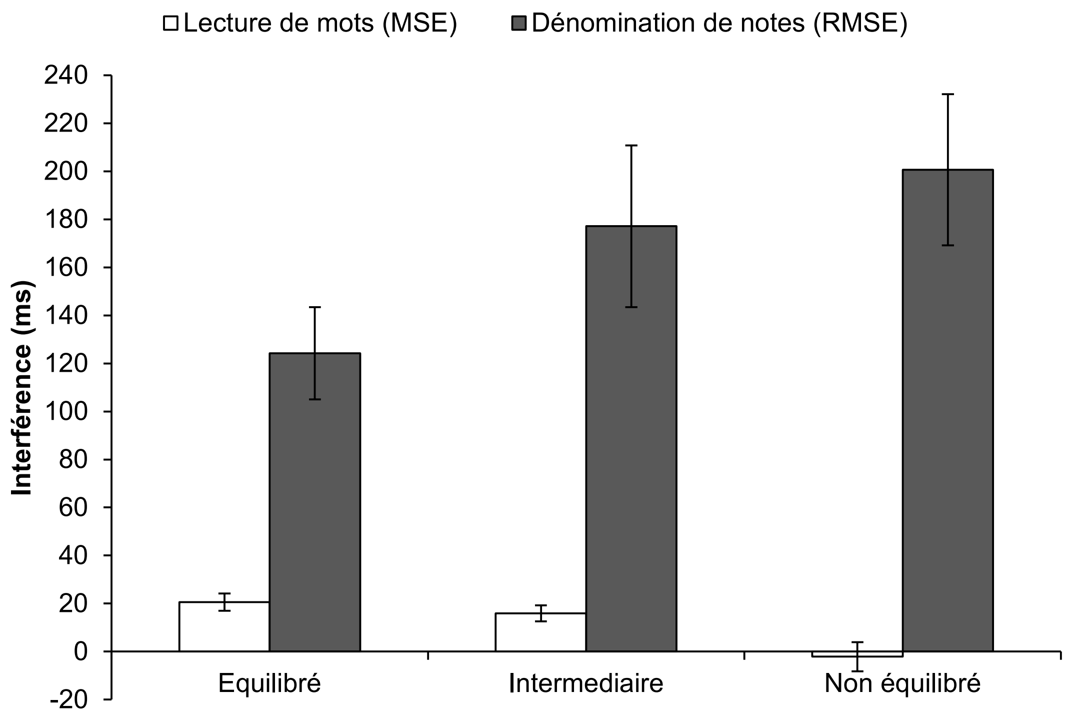

Figure 2. Quantité d'interférence (TRs en condition incongruente moins TRs en condition congruente) dans les tâches de lecture de mots et de dénomination de notes pour les trois groupes de sujets, rangés selon leurs performances aux tests d'habiletés en lecture et dénomination de notes. Groupe équilibré : la dénomination de notes était légèrement plus lente que la lecture de mots. Groupe non-équilibré : la dénomination de notes était largement plus lente que la lecture de mots ; Groupe intermédiaire : entre les deux groupes précédents. Les barres d'erreurs indiquent les erreurs types.

Le pattern observé quand la quantité d'interférence fut évaluée en fonction du pourcentage d'erreurs différait sensiblement de celui obtenu avec les TRs, mais cela résultait probablement du faible nombre d'erreurs recueilli dans les deux tâches. La corrélation entre la force relative et la quantité d'interférence était proche de 0 dans la tâche de lecture de mots, $r(20) = -.077$, $p = .733$, comme dans la tâche de dénomination de notes, $r(20) = .015$, $p = .946$.

Discussion

L'objectif principal de cette étude était d'observer si l'automaticité de la dénomination de notes persistait après un arrêt total et prolongé de la pratique musicale chez des sujets qui ont suivi une formation de solfège soutenue dans le passé. Les résultats préliminaires montraient clairement qu'un MSE était présent malgré une durée d'arrêt moyenne d'environ 14 ans. De plus, le MSE apparaissait chez 19 des 22 participants (à un niveau numérique).

Ainsi, un traitement qui a été hautement automatisé semble conserver une certaine irrépressibilité même s'il n'est plus pratiqué depuis de nombreuses années.

Nous avons également mis en évidence une corrélation négative entre le MSE et le RMSE. Une analyse plus approfondie indiquait que le pattern d'interférence dépendait de la force relative des deux traitements (lecture de mots et dénomination de notes) en compétition. En évaluant la force de chaque traitement par les TRs mesurés avec les tests d'habiletés, la force relative de la dénomination de notes et de la lecture de mots pouvait être obtenue en calculant la différence de TRs dans ces deux tâches. L'amplitude de cette différence déterminait alors la taille de l'interférence du MSE et du RMSE. En effet, nous avons constaté que la corrélation entre la force relative des deux traitements et la quantité d'interférence était négative dans la tâche de lecture de mots et positive dans la tâche de dénomination de notes. Nous avons ainsi répliqué les résultats obtenus dans l'Etude 2.

Nous envisageons toutefois de faire passer cette étude à un plus grand nombre de sujets afin de réaliser des analyses corrélationnelles plus abouties.

Bibliographie

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Howard University Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The Musical Stroop Effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 554–565.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126–135.

DISCUSSION GENERALE

Ce travail de thèse avait pour but d'examiner l'émergence et l'évolution du caractère obligatoire du traitement automatique. Pour satisfaire notre objectif, nous avons conçu une nouvelle situation expérimentale que nous avons appelée *Stroop musical*. Il s'agit d'une portée en clé de sol comprenant une note, présentée dans différentes positions, dans laquelle un nom de note est écrit. Notre première étude a permis de montrer, à l'aide de ce paradigme, que les musiciens traitaient plus lentement les noms de notes présentés dans une note incongruente, dans une tâche go/no-go (Expérience 1) et une tâche verbale (Expérience 2), par rapport à une situation où les noms de notes étaient congruents avec la position de la note (Grégoire et al., 2013). Nous avons donc mis en évidence un nouvel effet de type Stroop, que nous avons nommé *effet Stroop musical* (MSE), généré par l'automaticité de la dénomination de notes chez les musiciens.

Nous pourrions toutefois envisager que les musiciens se distinguent des non-musiciens par une plus grande sensibilité à l'interférence dans une situation de type Stroop (indépendamment de la nature des traitements en compétition). Mais de nombreux travaux suggèrent au contraire que les musiciens sont moins sensibles à l'interférence et ont un meilleur contrôle exécutif que les non-musiciens (e.g., Bialystok & DePape, 2009; Travis et al., 2011). Nous avons par ailleurs répliqué et renforcé le MSE dans notre seconde étude en utilisant une procédure simplifiée. Le MSE apparaît ainsi comme un effet robuste qui semble être la conséquence d'une pratique musicale intensive.

Cependant, une limite de notre paradigme est que nous ne disposons pas d'une condition neutre appropriée. Il nous est donc difficile de distinguer la facilitation et l'interférence. Nous avons réalisé des analyses à partir des données obtenues dans une condition où seuls le mot et la note étaient présentés. Les résultats indiquaient que le MSE proviendrait davantage de l'interférence générée par l'incongruence que par une éventuelle facilitation due à la congruence (Etudes 1 et 2). La validité de cette méthode reste néanmoins discutable. Il serait préférable d'utiliser une condition neutre convenable, mais la mise au point d'une ligne de base adéquate est un problème récurrent dans les tâches de type Stroop (voir, e.g., Melara & Algom, 2003; Neely, 1991; Sabri et al., 2001). Pour cette raison, certains auteurs ont suggéré de ne pas en utiliser (Jonides & Mack, 1984; Lindsay & Jacoby, 1994).

Il est communément admis que l'interférence provient du caractère obligatoire de la lecture des mots dans une tâche de type Stroop (MacLeod, 1991). En revanche, certains travaux ont supposé que la facilitation ne résulterait pas du même processus sous-jacent (e.g., MacLeod, 1998). La condition congruente serait en effet contaminée par des inversions de tâche occasionnelles (i.e., le sujet lirait le mot au lieu de dénommer la couleur) qui réduiraient le TR moyen[22] (Kane & Engle, 2003; MacLeod & MacDonald, 2000). Mais l'hypothèse de lecture des mots par inadvertance a été récusée par une étude récente (Roelofs, 2010). En utilisant la tâche de Stroop couleur-mot avec des sujets bilingues, les inversions de tâche sont facilement décelables (dans la tâche de dénomination de couleur) si le langage de réponse est différent du langage des mots écrits. Roelofs tire parti de cette situation pour montrer que la facilitation provient de la convergence des informations conflictuelles. Il en conclut que l'interférence et la facilitation résultent du même processus sous-jacent (p. 420).

Si le MSE représente un effet conjoint de l'interférence et de la facilitation (même si cette dernière est minime), il apparaît donc raisonnable de considérer que ces deux effets proviennent du traitement automatique de la dimension non-pertinente. Néanmoins, dans la tâche de lecture de mots (du paradigme de Stroop musical), la condition congruente pourrait entraîner, de manière occasionnelle, des inversions de tâche. Mais cela ne devrait pas conduire à une surestimation du MSE. En effet, nos données indiquent que la lecture de mots est plus rapide et certainement plus automatisée que la dénomination de notes (Etude 2). Une inversion de tâche dans la condition congruente devrait donc *amplifier* le TR et ainsi réduire le MSE.

Par conséquent, l'absence d'une condition neutre n'apparaît pas comme un problème majeur dans notre situation. Nous n'avons donc pas utilisé de ligne de base dans nos études développementales (Etudes 3 et 4), de même que dans notre dernière investigation (Etude 5). Le recours à cette procédure ne semble pas rédhibitoire pour répondre aux questions que nous avons soulevées. En effet, nous souhaitions évaluer le caractère obligatoire du traitement automatique de la dénomination de notes chez les musiciens, principalement son évolution au cours de la pratique, et le MSE représente un indicateur de ce critère.

L'évolution linéaire du caractère obligatoire en fonction de la pratique

Le paradigme de Stroop musical ouvre donc la voie à de nouvelles perspectives pour étudier les automatismes. Tout d'abord, il peut permettre de mesurer l'évolution du MSE en fonction du niveau de pratique musicale en évitant les biais liés à l'âge des sujets. En effet, la plupart des travaux qui se sont penchés sur cette question en utilisant une tâche de type Stroop comportent un problème majeur (et difficile à contourner pour des raisons éthiques et pratiques) : l'âge des sujets est largement confondu avec le degré d'automaticité de la

[22] Les inversions de tâche dans la condition congruente ne sont pas repérables (dans la tâche de dénomination de couleur) car l'information à ignorer (i.e., le mot) est similaire à l'information pertinente (i.e., la couleur).

dimension interférente (principalement, la lecture). Ainsi, l'évolution de l'interférence peut aussi bien être expliquée par les habiletés à traiter la dimension non-pertinente que par des facteurs liés à l'âge des participants, comme le niveau de maturation neurale et cognitive.

La pratique musicale peut être plus facilement découplée de l'âge que la lecture, car l'apprentissage du solfège n'est pas aussi régulé que l'éducation scolaire. Nous avons donc exploité cette opportunité en testant plusieurs groupes de sujets, d'un âge sensiblement équivalent, dont le niveau de solfège variait de 1 à 3 ans (Etude 3, Expérience 1), et de 1 à 5 ans (Etude 3, Expérience 2). Nos résultats indiquent que le MSE augmente linéairement en fonction de la pratique musicale. Des analyses corrélationnelles montrent que cet effet est indépendant de l'âge des participants. Par ailleurs, nous constatons une tendance linéaire négative des TRs en fonction du niveau de solfège dans la tâche de Stroop musical.[23] L'augmentation du MSE ne peut donc pas être expliquée par un ralentissement général des TRs.

Nous n'avons pas répliqué la relation linéaire positive entre le MSE et le niveau de pratique musicale dans l'Etude 4. Une augmentation du MSE était observée à un niveau descriptif, mais cet effet n'était pas significatif. Cependant, l'objectif de cette étude n'était pas le même que dans l'Etude 3, et l'âge des sujets n'a pas été maintenu constant. Il est donc possible que des effets liés à l'âge des participants aient contribué à atténuer le MSE.

Ainsi, le caractère obligatoire de la dénomination de notes semble augmenter de façon croissante et monotone avec la pratique, ce qui nous incline à penser que la décroissance observée dans les études qui utilisent le paradigme de Stroop couleur-mot (ou image-mot) serait liée à l'âge des sujets, notamment à une plus grande capacité à inhiber la dimension non-pertinente. Néanmoins, on pourrait supposer que nos données ne sont pas suffisamment exhaustives pour exclure la possibilité de mettre en évidence une relation en U inversé. Une récupération du contrôle (voir Logan, 1985; Tzelgov et al., 1990) serait éventuellement susceptible d'intervenir après la cinquième année de solfège, avec un plus grand degré d'expertise de la dénomination de notes. Mais nous avons plusieurs raisons de penser que cette hypothèse est hautement improbable. Tout d'abord, la décroissance de l'interférence obtenue dans les tâches de Stroop couleur-mot (ou image-mot) apparaît, la plupart du temps,

[23] Ce résultat pourrait provenir de plusieurs effets conjoints qui se développeraient au cours de la pratique : un temps de dénomination de notes plus rapide (nous avons observé une tendance linéaire négative des TRs dans le test d'habiletés en dénomination de notes), un coût cognitif moins important pour traiter les notes (ce qui libérerait des ressources attentionnelles pour traiter plus rapidement l'information verbale), ou encore une plus grande familiarité avec les stimuli musicaux (les non-musiciens étaient plus lents que les musiciens en première année de solfège dans la tâche de Stroop musical).

relativement tôt au cours de l'apprentissage (en général, entre la deuxième et la troisième année de pratique de la lecture; e.g., Dash & Dash, 1982; Peru et al., 2006; Schiller, 1966). Il serait surprenant que cette décroissance intervienne après la cinquième année de solfège dans notre situation, car le système de notation musicale semble nettement moins complexe que celui de la lecture, ce qui suggère que l'automaticité de la dénomination des notes devrait se développer plus rapidement que l'automaticité de la lecture. On pourrait toutefois arguer que les notes de musique ne sont pas omniprésentes dans notre environnement, contrairement aux mots, mais les enfants sont testés uniquement avec les notes qui constituent la première octave de la clé de sol, c'est-à-dire les notes les plus fréquemment rencontrées en cours de solfège. De plus, il apparaît que l'amplitude du MSE chez des adultes est plus forte que chez les enfants du groupe 5 de l'Etude 3 (Expérience 2), dans une situation pourtant moins favorable. La proportion d'items congruents est plus faible (33%) que dans l'Etude 3 (50%), une situation connue pour générer moins d'interférence (Lowe & Mitterer, 1982). Par ailleurs, les capacités d'inhibition sont théoriquement plus importantes chez les adultes (voir par exemple, Bédard et al., 2002). Ce résultat semble donc indiquer que le MSE continue d'augmenter au-delà de la cinquième année de solfège.

Plusieurs études avaient déjà mentionné, avec des sujets d'un âge équivalent, que l'interférence amplifiait lorsque le niveau de pratique de la dimension à ignorer dans une tâche de type Stroop s'accentuait. Mais la portée de ces travaux semblait relativement réduite. Les résultats étaient basés sur l'observation de deux groupes de sujets seulement (Martin, 1978 ; Stanovich et al., 1981), ou alors, il s'agissait d'une situation d'apprentissage artificielle en laboratoire qui se déroulait sur une période relativement courte (20 semaines) avec un petit échantillon (4 sujets ; Macleod & Dunbar, 1988). Une étude plus récente révélait que le niveau d'habiletés du second langage chez des sujets bilingues était corrélé positivement avec l'amplitude de l'effet Stroop (Braet et al., 2011), mais il n'existe pas de consensus, à l'heure actuelle, concernant la nature des processus en cours d'automatisation lors de l'apprentissage d'une seconde langue.

Néanmoins, l'ensemble de ces résultats converge vers une conclusion analogue à la nôtre : le caractère obligatoire du traitement automatique tend à augmenter de manière linéaire avec la pratique quand l'âge est maintenu constant. Il semblerait donc que les différentes propriétés communément attribuées aux automatismes évoluent en parallèle. Cette question est débattue depuis longtemps dans la littérature (e.g., Johnson & Dark, 1986 ; Paap & Ogden, 1981). Il est largement admis que l'efficacité d'un traitement s'accentue de façon monotone en fonction du niveau de pratique (e.g., Lien et al., 2006), de même que sa vitesse d'exécution (e.g., Crossman, 1959). Sur la base de nos résultats, nous pouvons conclure que le caractère obligatoire du traitement automatique évolue de la même manière.

La force relative des deux traitements en compétition détermine l'amplitude du MSE et du RMSE

Le paradigme de Stroop musical offre également la possibilité de confronter deux traitements hautement automatisés après une pratique intensive et naturelle : la lecture de mots et la dénomination de notes. Nous avons exposé des musiciens aux situations congruentes et incongruentes de notre paradigme en leur demandant, soit de lire le mot en ignorant la position de la note sur la portée, soit de dénommer la note sans tenir compte du mot écrit à l'intérieur. Nous souhaitions ainsi observer si un effet Stroop se manifesterait dans les deux tâches conflictuelles.

L'asymétrie de l'effet Stroop, selon laquelle les mots interfèrent sur la dénomination de la couleur alors que la couleur n'affecte pas la lecture des mots, est un résultat bien documenté. De manière générale, peu d'études ont mis en évidence un effet Stroop inversé (i.e., une altération de la lecture de mots par une dimension concurrente) dans des conditions naturelles de traitement. Plusieurs travaux suggèrent qu'un effet Stroop inversé pourrait être obtenu, mais au détriment de l'effet Stroop lui-même, comme si l'interférence était intrinsèquement unidirectionnelle (e.g., Akiva-Kabiri & Henik, 2012 ; Palef & Olson, 1975). L'observation des effets Stroop bidirectionnels reste rare et fugace (e.g., Melara & Mounts, 1993), et elle pourrait être due à un mélange d'effets unidirectionnels qui apparaitraient dans des directions opposées pour les différents participants.

Les résultats observés dans notre seconde étude vont clairement à l'encontre de l'hypothèse d'une unidirectionnalité de l'interférence Stroop. En effet, nous avons répliqué le MSE, mais nous avons également montré que la dénomination d'une note était plus lente dans la condition incongruente que dans la condition congruente, ce qui témoigne d'un MSE inversé (RMSE). Le MSE et le RMSE étaient obtenus dans des conditions strictement identiques, qui correspondent autant que possible aux conditions de lecture et de dénomination de notes rencontrées dans des situations naturelles. Ainsi, nos données rendent compte d'une interférence mutuelle entre deux processus en compétition. De plus, 24 des 26 musiciens testés montraient les deux effets sur les TRs (à un niveau numérique), ce qui exclut la possibilité que les effets Stroop bidirectionnels proviennent d'un artefact dû à des statistiques de groupes.

La co-occurrence du MSE et du RMSE semble remettre en cause la validité du modèle connexionniste de Cohen et al. (1990) selon lequel l'automaticité dépend de la force d'un traitement, qui est elle-même liée à la quantité de pratique. Cohen et al. considèrent que la force relative, comparée à un traitement concurrent, détermine si un processus produira ou sera sujet à l'interférence dans une tâche de type Stroop (p. 348). En conséquence, un processus A ne peut pas être à la fois fortement interférent sur un processus B (cela impliquerait A > B) et fortement interféré par un processus B (cela impliquerait B < A), et réciproquement. Des effets Stroop bidirectionnels ne pourraient être obtenus que si la force des deux dimensions en compétition était similaire. Cependant, le modèle échoue à simuler une interférence mutuelle quand les deux processus ont une force comparable (les effets sont négligeables, p. 349).

Néanmoins, des analyses supplémentaires montraient que nos données étaient au contraire conformes au modèle de Cohen et al. (1990) si l'on considérait que la force relative, comparée avec un traitement concurrent, déterminait la propension relative du processus le plus fort, à interférer sur un processus plus faible, et à être interféré par ce processus plus faible. En évaluant la force des traitements (la lecture de mots et la dénomination de notes) par la vitesse mesurée dans les tests d'habiletés, nous avons constaté que la force de la lecture était beaucoup plus importante que la force de la dénomination de notes. De ce fait, la lecture de mots provoquait une plus grande interférence sur la dénomination de notes que l'inverse : le RMSE était nettement plus fort que le MSE. Mais l'analyse des différences individuelles rendait encore mieux compte de notre hypothèse. Nous avons estimé la force relative des deux traitements par la différence des TRs entre les deux tests d'habiletés (dénomination de notes - lecture de mots). Cette mesure était corrélée négativement avec le MSE et positivement avec le RMSE. Nous avons alors constitué trois sous-groupes de sujets en fonction de l'amplitude des différences obtenues en calculant la force relative. Le sous-groupe de musiciens qui montraient des performances relativement équivalentes dans les deux tests (groupe équilibré) présentait l'interférence mutuelle la plus homogène, et le MSE n'était pas significativement différent du RMSE. A l'opposé, le sous-groupe de sujets qui exhibait le plus fort déséquilibre entre les deux traitements manifestait la plus forte asymétrie entre les deux effets, avec un MSE qui ne différait pas statistiquement de la chance. Ainsi, le pattern d'interférence dépendait de la force relative des deux traitements en compétition.

Nous avons répliqué ces résultats dans une étude qui avait pour but d'évaluer l'évolution concurrente du MSE et du RMSE avec des enfants en cours d'apprentissage des deux dimensions : la lecture et la dénomination de notes (Etude 4). Les sujets ont réalisé les mêmes tâches que dans l'Etude 2 (la procédure a toutefois été simplifiée). Les résultats révélaient que le MSE et le RMSE étaient présents conjointement dans les cinq groupes de participants. De plus, l'amplitude des deux effets était liée à la force relative des deux traitements. Notre dernière étude reproduit également le même pattern de résultats avec des adultes qui ont cessé toute pratique musicale après avoir suivi une formation de solfège soutenue dans le passé (Etude 5).

Nous suggérons que la différence entre les résultats collectés avec le paradigme de Stroop couleur-mot (ou image-mot) et ceux obtenus avec notre paradigme note-mot chez les musiciens révèle l'importance de la contingence stimulus-réponse dans la formation des automatismes et leur pouvoir interférent. La contingence entre les événements a un rôle essentiel dans le renforcement des relations associatives (Perruchet & Poulin-Charronnat, 2012 ; Rescorla, 1968). Le niveau de contingence indique notamment comment un événement est prédictif d'un autre événement. Par exemple, la contingence entre la perception d'un objet et sa dénomination est relativement faible dans la vie quotidienne. En effet, la présence d'un objet dans notre champ visuel n'implique pas l'évocation de son nom (même de façon subvocale). Un raisonnement analogue peut être fait pour la dénomination de couleur. Nous n'avons, la plupart du temps, aucune raison de dénommer la couleur d'un objet que nous appréhendons dans notre environnement. Au contraire, la contingence entre la perception d'un mot écrit et sa production orale ou subvocale est considérablement plus forte. De même, la contingence entre la perception d'une note sur une portée et sa dénomination doit être très élevée. Pour un musicien, une note n'est pratiquement jamais présente dans l'environnement en dehors du contexte dans lequel elle doit être nommée. Ainsi, l'automatisation de la dénomination de notes serait favorisée par une forte contingence stimulus-réponse. La pratique musicale permettrait, de cette manière, d'atteindre un niveau d'automaticité en dénomination de notes suffisamment élevé pour interférer avec la lecture de mots.

Les résultats des Etudes 2, 4, et 5 semblent montrer que les difficultés rencontrées dans la littérature pour mettre en évidence un effet Stroop inversé sont liées davantage au caractère non-automatique du traitement interférent qu'à l'automaticité de la lecture. L'asymétrie habituellement observée dans les tâches de type Stroop provient donc probablement d'une différence trop marquée entre la force relative de la lecture et de l'autre dimension en compétition (e.g., la couleur ou la dénomination d'images). Le paradigme de Stroop musical permet de mettre en concurrence deux traitements qui possèdent un certain degré d'automaticité. Ainsi, dans cette situation, les deux effets, le RSE et le RMSE, peuvent être observés simultanément. Nos données démontrent par conséquent que l'extrême automaticité de la lecture ne l'immunise pas contre l'interférence.

Nous envisageons de vérifier si la force relative entre les deux dimensions concurrentes pourrait rendre compte des effets obtenus avec la tâche de Stroop couleur-mot en s'intéressant à des enfants, d'un âge donné, qui commencent à apprendre la lecture (en cours primaire ou élémentaire). Le niveau de pratique des deux traitements en compétition (i.e., la lecture de mots et la dénomination de couleur) devrait être relativement équivalent à cette période du développement. Peru et al. (2006) ont d'ailleurs montré qu'une interférence était présente conjointement dans la tâche de dénomination de couleur et dans la tâche de lecture de mots entre 7 et 9 ans. Mais ces auteurs ne présentent pas d'analyses individuelles qui pourraient nous permettre d'éprouver notre hypothèse. De futures investigations devraient parvenir à nous éclairer sur cette question.

L'automaticité de la dénomination de notes subsiste à l'arrêt total et prolongé de la pratique

Le paradigme de Stroop musical permet également d'explorer un aspect des automatismes qu'il est inconcevable d'étudier avec d'autres tâches de type Stroop. Nous pouvons en effet déterminer si le traitement automatique de la dénomination de notes persiste après un arrêt total et prolongé de la pratique. La possibilité d'évaluer le devenir de processus qui ont été automatisés dans le passé, après avoir cessé toute activité, apparaît relativement complexe avec des apprentissages naturels. Cela semble même parfaitement illusoire à propos de la lecture étant donné l'ubiquité des mots dans notre environnement quotidien. Les rares études qui se sont penchées sur la question concernaient des apprentissages en laboratoire, et les périodes d'arrêt étaient relativement succinctes, entre 3 mois (MacLeod & Dunbar, 1988) et un an (Kolers, 1976).

Nous nous sommes intéressés à des sujets adultes qui ont complètement délaissé la pratique musicale depuis au moins trois ans après avoir reçu une formation de solfège soutenue. Les résultats préliminaires indiquaient un MSE significatif malgré une durée d'arrêt moyenne d'environ 14 ans. De plus, cet effet se manifestait chez 19 des 22 participants (à un niveau numérique). Il apparaît ainsi qu'un traitement qui a été hautement automatisé dans le passé semble conserver son caractère irrépressible même s'il n'est plus pratiqué depuis de nombreuses années.

Vers une confrontation des deux grandes conceptions de l'automaticité

Une autre possibilité offerte à l'exploration expérimentale par le paradigme de Stroop musical concerne la confrontation entre les deux grandes conceptions de l'automaticité : le renforcement d'une procédure algorithmique (e.g., Anderson, 1992; Rosenbloom & Newell, 1986) et la récupération directe en mémoire (Logan, 1988). Nous pourrions en effet évaluer quel type d'apprentissage de la dénomination de notes est le plus bénéfique. Est-ce qu'un apprentissage absolu, dans lequel chaque note est apprise pour elle-même sans se référer à une note préalablement apprise (similaire à un processus de récupération immédiate en mémoire) serait, comme le prédirait la théorie de Logan (1988), plus bénéfique qu'un apprentissage « relatif » s'appuyant sur des notes repères (similaire à un apprentissage de type algorithmique) ? Nous pourrions envisager, par exemple, de tirer parti des différences de méthodes d'apprentissage entre les écoles de musique. Des sujets provenant d'une école qui enseigne la lecture de partition à partir de notes repères auraient-ils un MSE plus faible que des participants (du même âge et du même niveau de solfège) qui ont suivi un apprentissage absolu dans une autre école de musique ? Il serait également possible de contrôler directement la méthode d'apprentissage avec des groupes d'adultes volontaires pour apprendre le solfège. Les résultats de cette étude pourraient avoir des répercutions pédagogiques importantes sur les méthodes d'apprentissage de la dénomination de notes dans les écoles de musique.

BIBLIOGRAPHIE

Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A developmental fMRI study of the stroop color-word task. *Neuroimage, 16*, 61–75.

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical stroop effect in absolute pitch possessors. *Experimental Psychology, 59*, 272–278.

Akiva-Kabiri, L. & Henik, A. (2014). Additional insights: Commentary on “ The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269–278). *Experimental Psychology, 61*, 75–77.

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36*, 550–568.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Howard University Press.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105*, 165–180.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355–365.

Armengol, C. G. (2002). Stroop test in Spanish: children’s norms. *The Clinical Neuropsychologist, 16*, 67–80.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology, 105*, 181–199.

Bédard, A. C., Nichols, S., Barbosa, J. A., Schachar, R., Logan, G. D., & Tannock, R. (2002). The development of selective inhibitory control across the life span. *Developmental Neuropsychology, 21*, 93–111.

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 565–574.

Birnbaum, I. M., Taylor, T. H., Johnson, M. K., & Raye, C. L. (1987). Is event frequency encoded automatically? The case of alcohol intoxication. *Journal of Experimental*

Psychology: Learning, Memory and Cognition, 13, 251–258.

Blais, C., & Besner, D. (2006). Reverse stroop effects with untranslated responses. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1345–1353.

Blais, C., & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review, 14*, 466–469.

Braet, W., Noppe, N., Wagemans, J., & Op de Beeck, H. P. (2011). Increased Stroop interference with better second-language reading skill. *Quarterly Journal of Experimental Psychology, 64*, 596–607.

Broadbent, D. (1958). *Perception and communication.* London: Pergamon Press.

Carlson, R. A., & Lundy, D. H. (1992). Consistency and restructuring in cognitive procedural sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 127–141.

Carter, C. S., Mintun, M., Cohen, J. D. (1995). Interference and facilitation effects during selective attention: an $H_2{}^{15}O$ PET study of Stroop task performance. *NeuroImage, 2*, 264–272.

Catling, J. C., Dent, K., Johnston, R. A., & Balding, R. (2010). Age of acquisition, word frequency, and picture-word interference. *The Quarterly Journal of Experimental Psychology, 63*, 1304–1317.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11*, 63–65.

Chen, H. C, & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 397–401.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.

Comalli, P. E., Jr., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology, 100*, 47–53.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed skills. *Ergonomics, 2*, 153–166.

Dash, J., & Dash, A. S. (1982). Cognitive developmental studies of the Stroop phenomena: Cross-sectional and longitudinal data. *Indian Psychologist, 1*, 24–33.

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Neuropsychology, Development, and Cognition, Section B: Aging, Neuropsychology and Cognition, 10*, 85–98.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80–90.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 470–484.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 622–639.

Durgin, F. H. (2003). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics, 65*, 367–378.

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior, 10*, 297–302.

Ehri, L. C. (1976). Do words really interfere in naming pictures? *Child Development, 47*, 502–505.

Ehri, L. C., & Wilce, L. S. (1979). Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology, 27*, 352–364.

Ellis, N. R., Woodley-Zanthos, P., Dulaney, C. L., & Palmer, R. L. (1989). Automatic-effortful processing and cognitive inertia in persons with mental retardation. *American Journal on Mental Retardation, 93*, 412–423.

Everatt, J., Warner, J., & Miles, T. R. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia, 3*, 222–228.

Faccioli, C., Peru, A., Rubini, E., & Tassinari, G. (2008). Poor readers but compelled to read: Stroop effects in developmental dyslexia. *Child Neuropsychology, 14*, 277–283.

Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in "ignoring" irrelevant information. *Memory & Cognition, 7*, 86–94.

Francolini, C. M., & Egeth, H. (1980). On the automaticity of “automatic” activation: Evidence of selective seeing. *Perception & Psychophysics, 27*, 331–342.

Gast, A. (2014). What is learned, and when? Commentary on “ The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269–278). *Experimental Psychology, 61*, 71–74.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640–654.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 875–894.

Girelli, L., Lucangeli, D., Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology, 76*, 104–122.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014). Is the musical Stroop effect able to keep its promises? A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology, 61*, 80–83.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (in press). About the unidirectionality of interference: Insight from the musical Stroop effect. *Quarterly Journal of Experimental Psychology.*

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (submitted). How does Stroop interference vary with practice? A reappraisal from the musical Stroop paradigm.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (in preparation). How are two automatisms developing concurrently with practice? Pitting note naming against reading in musicians.

Guttentag, R. E., & Haith, M. M. (1978). Automatic processing as a function of age and reading ability. *Child Development, 49*, 707–716.

Guttentag, R. E., & Haith, M. M. (1979). A developmental study of automatic word processing in a picture classification task. *Child Development, 50*, 894–896.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.

Heine, A., Tamm, S., De Smedt, B., Schneider, M., Thaler, V., Torbeyns, J., et al. (2010). The numerical stroop effect in primary school children: A comparison of low, normal, and high achievers. *Child Neuropsychology, 16*, 461–477.

Helland, T., & Asbjørnsen, A. (2000). Executive functions in dyslexia. *Child Neyropsychology, 6*, 37–48.

Hicks, C., & Jackson, P. (1981). A study of the relationship between the Stroop effect and reading age in dyslexic subjects. *Journal of Research in Reading, 4*, 29–33.

Hodges, D., & Nolker, B. (2011). The acquisition of music reading skills. In R. Colwell & P. Webster (Eds.), *MENC handbook of research on music learning, volume II: applications* (pp. 61–91). Oxford, UK: Oxford University Press.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic and intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jastrow, J. (1906). *The subconscious.* Boston: Houghton Mifflin.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. Annual Review of Psychology, 37, 43–75.

Jonides, J., & Mack, R. (1984). On the costs and benefits of costs and benefits. *Psychological Bulletin, 96*, 29–44.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D., & Chajczyk, D. (1983). Test of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 497–509.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 181–212). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). Orlando, FL: Academic Press.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop Interference. *Journal of Experimental Psychology: General, 132,* 47–70.

Kapoula, Z., Lê, T. T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., Quilicci, C., & Yang, Q. (2010). Poor Stroop performances in 15-year-old dyslexic teenagers. *Experimental Brain Research, 203*, 419–425.

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 137*, 1445–1452.

Klenberg, L., Korkman, M., Lahti Nuuttila, P. (2001). Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Developmental Neuropsychology, 20*, 407–428.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 554–565.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for symmetric connections between bilingual memory representations. *Journal of Memory and Language, 39*, 149–174.

Laberge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Lee, T. M., & Chan, C. C. (2000). Stroop interference in Chinese and English. *Journal of Clinical and Experimental Neuropsychology, 22*, 465–471.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Lien, M. C., Allen, P. A., Ruthruff, E., Grabbe, J., McCann, R. S., & Remington, R. W. (2006). Visual word recognition without central attention: Evidence for greater automaticity with advancing age. *Psychology and Aging, 21*, 431–447.

Lindsay, D. S., & Jacoby, L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception & Performance, 20*, 219–234.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523–553.

Logan, G. D. (1982). On the ability to inhibit complex movements: A Stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 778–792.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39*, 367–386.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Logan, G. D. (1989). Automaticity and cognitive control. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 52–74). New York: Guilford Press.

Logan, G. D. (1991). Automaticity and memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 347–366). Hillsdale, NJ: Erlbaum.

Logan, G. D. (1992). Attention and preattention in theories of automaticity. *American Journal of Psychology, 105*, 317–339.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading and Writing Quarterly, 13*, 123–146.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is Extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 179–195.

Lowe, D. G., & Mitterer, J. O. (1982). Selective and divided Attention in a Stroop task. *Canadian Journal of Psychology, 36*, 684–700.

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the Color Stroop Test: A comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics, 51*, 135-142.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7*, 485–495.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*, 201–211.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126–135.

MacLeod, C. M., & MacDonald, P. A. (2000). Inter-dimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383–391.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 304–315.

Martin, M. (1978). Speech recoding in silent reading. *Memory & Cognition, 6*, 108–114.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*, 422–471.

Melara, R. D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21*, 627–645.

Moeller, B., & Frings, C. (2014). How automatic is the musical the musical Stroop effect ? Commentary on " The musical Stroop effect: Opening a new avenue to research on automatisms" by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269–278). *Experimental Psychology, 61*, 68–70.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.

Nealis, P. M. (1973). The Stroop phenomenon: Some critical tests of the response competition hypothesis. *Perceptual and Motor Skills, 37*, 147–153.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88*, 1–15.

Ogura, C. (1980). Formation of learning-set in the task of Stroop color-word test. *Journal of Child Development, 16*, 30-36.

Pace, A. J., & Golinkoff, R. M. (1976). Relationship between word difficulty and access of single-word meaning by skilled and less skilled readers. *Journal of Educational Psychology, 58*, 760–767.

Paap, K. R., & Ogden, W. C. (1981). Letter encoding is an obligatory but capacity-demanding operation. *Journal of Experimental Psychology : Human Perception and Performance, 7*, 518–527.

Palef, S. R., & Olson, D. R. (1975). Spatial and verbal rivalry in a Stroop-like task. *Canadian Journal of Psychology, 29*, 201–209.

Pascal, B. (1954). Œuvres complètes. Paris: Gallimard, La Pléiade.

Perruchet, P. (1988). Les automatismes cognitifs. Liège : Mardaga.

Perruchet, P., & Poulin-Charronnat, B. (2012). Word segmentation: Trading the (new, but poor) concept of statistical computation for the (old, but richer) associative approach. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 119–143). Berlin: De Gruyter Mouton.

Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes de Biologie, 144*, 45–62.

Peterson, J., Lanier, L. H., & Walker, H. M. (1925). Comparisons of white and negro children in certain ingenuity and speed tests. *Journal of Comparative Psychology, 5*, 271–283.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 153–175). Hillsdale, NJ: Erlbaum.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology, 54*, 251–282.

Quantz, J. O. (1897). Problems in the psychology of reading. *Psychological Review Monographs, 2*(No. 5), 1–51.

Rand, G., Wapner, S., Werner, H., & McFarland, J. H. (1963). Age differences in performance on the Stroop color-word test. *Journal of Personality, 31*, 534–558.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446–461.

Reason, J. (1984). Lapses of attention in everyday life. In R. Parasuraman & D.R. Davies (Eds), *Varieties of attention*, Orlando, F.L., Academic Press.

Regan, J. E. (1981). Automaticity and learning: Effects of familiarity on naming letters. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 180–195.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66*, 1–5.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249–284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*, 88–125.

Roelofs, A. (2005). The visual-auditory color-word stroop asymmetry and its time course. *Memory & Cognition, 33*, 1325–1336.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 353–368.

Roelofs, A. (2010). Attention and facilitation: Converging information versus inadvertent reading in Stroop task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 411–422.

Roelofs, A., & Hagoort, P. (2002). Control of language use: Cognitive modeling of the hemodynamics of Stroop task performance. *Cognitive Brain Research, 15*, 85–97.

Rosenbloom, P. S., & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michaliski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2, pp. 247–288). Los Altos, CA: Morgan Kaufmann.

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development, 48*, 643–647.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46*, 247–253.

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs nonsymbolic number magnitude processing. *Cognition, 102*, 361–395.

Ruthruff, E., Allen, P. A., Lien, M. C., & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychonomic Bulletin and Review, 15*, 337–343.

Sabri, M., Melara, R. D., Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. Journal of Experimental Psychology. *Human Perception and Performance, 27*, 515–537.

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9*, 132–141.

Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology, 72*, 105–108.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1984). Automatic and controlled processing and attention. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 1–17). New York: Academic press.

Schneider, W., & Fisk, A. D. (1984). Automatic category search and its transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 1–15.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schulz, T. (1979). Components of the reaction time Stroop-task. *Psychological Research, 40*, 377–395.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127–190.

Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology, 93*, 281–288.

Smith, E. R., & Lerner, M. (1986). Development of automatism of social judgements. *Journal of Personality and Social Psychology, 50*, 246–259.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109*, 373–392.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4*, 215–230.

Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10*, 72–100.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior, 13*, 57–74.

Stewart, L., Walsh, & Frith, U. (2004). Reading music modifies spatial mappings in pianists. *Perception & Psychophysics, 66*, 183–195.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Sumiya, H., & Healy, A. F. (2008). The Stroop effect in English–Japanese bilinguals: The effect of phonological similarity. *Experimental Psychology, 55*, 93–101.

Travis, F., Harung, H. S., & Lagrosen, Y. (2011). Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians: Interpreted in light of a Unified Theory of Performance. *Consciousness and Cognition, 20*, 1256–1264.

Tzelgov, J. (1997). Automatic but conscious: That is how we act most of the time. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 217–230). Mahwah, NJ: Erlbaum.

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 760–771.

Tzelgov, J., Henik, A., Sneg, R., & Baruch, B. (1996). Unintentional word reading via the phonological route: The Stroop effect with cross-script homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 336–349.

Tzelgov, J., & Kadosh, R. C. (2009). From automaticity to control in bilinguals. *Trends in Cognitive Sciences, 13*, 444–452.

Tzelgov, J., Yehene, V., Kotler, L., & Alon, A. (2000). Automatic comparisons of artificial digits never compared: Learning linear ordering relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 103–120.

Tzelgov, J., Yehene, V., & Naveh-Benjamin, M. (1997). From automaticity to memory and vice versa: On the relations between automaticity and memory. In J. Brzezinski, B. Krause, & T. Maryszewki (Eds.), *Idealization in psychology: Poznan studies of sciences and the humanities* (Vol. 55, pp. 239–261). Amsterdam: Rodopi.

Verdonschot, R. G., La Heij, W., & Schiller, N. O. (2010). Semantic context effects when naming Japanese kanji, but not Chinese hànzì. *Cognition, 115*, 512–518.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging, 13*, 120–126.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition, 13*, 304–319.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*, 158–177.

Wundt, W. (1897). *Outlines of psychology* (C. H. Judd, Trans.). Leipzig, Germany: Wilhelm Engelmann. (Original work published 1896)

Wundt, W. (1903). *Grundzüge der physiologischen Psychologie* [Principles of physiological psychology] (Vol. 3). Leipzig, Germany: Wilhelm Engelmann.

Zakay, D. (2014). Can the ‘‘musical Stroop’’ task replace the classical Stroop task? Commentary on “ The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (*Experimental Psychology,* 2013, Vol. 60, pp. 269–278). *Experimental Psychology, 61*, 78–79.

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology, 39*, 414–423.

Zysset, S., Müller, K., Lohmann, G., von Cramon, D. Y. (2001). Color-word matching Stroop task: separating interference and response conflict. *NeuroImage, 13*, 29–36.

ANNEXES

Annexe 1

Additional Insights: Commentary on Grégoire, Perruchet and Poulin-Charronnat (2013)
The Musical Stroop Effect: Opening a New Avenue to Research on Automatisms

Lilach Akiva-Kabiri & Avishai Henik

Abstract

In their paper "The Musical Stroop Effect: Opening a New Avenue to Research on Automatisms", Grégoire, Perruchet and Poulin-Charronnat (2013) use a musical Stroop-like task to demonstrate the automaticity of musical note naming in musicians. In addition, the authors suggest that music training can serve as a tool in order to study the acquisition of automaticity. In the following commentary, we aim to address three main issues concerning the paper by Grégoire et al. (2013). First, we will suggest some additional interpretations of the results; specifically, we will relate to the association between music and space. Second, we will discuss a methodological issue dealing with interference, facilitation and the role of the neutral condition. We suggest that the study by Grégoire et al. (2013) lacks a proper neutral condition and thus it is impossible to assert that the congruency effect is interference based. Third, we will discuss the authors' suggestion of using the musical Stroop effect as a tool for studying automatism. We consider the practical relevance of music training as a tool for studying the acquisition of automaticity by pointing out that music training is highly heterogeneous.

In their paper, Grégoire, Perruchet and Poulin-Charronnat (2013) study the automaticity of musical note naming. The authors used a go-no-go task and a Stroop-like task in two separate experiments in order to demonstrate the effect of incongruent musical notation on reading musical note names in trained musicians and controls. In separate experiments, participants were asked to respond to a musical note name and refrain from responding to a regular word (e.g., je (in French); Experiment 1) or to read a written word or note name (e.g., la; Experiment 2). When the note name was congruent with its spatial position on the staff, performance was faster, compared with the incongruent condition, for musicians but not for non-musician controls. The authors suggested that only musicians were affected by a note's spatial position on the staff. Moreover, they suggested musical training can serve as a tool for studying the acquisition of automatization because it enables manipulating practice.

We find the results of Grégoire et al. (2013) very interesting mostly for of two reasons: first, because of the intriguing association of musical tones to the spatial position on the staff; and second, because, unlike the original Stroop effect (1935) and its following replications (for a review see MacLeod, 1991), the authors found a compatibility effect when the irrelevant dimension was the spatial position on the musical staff and the relevant dimension was word meaning. Thus, the current task/results are similar to a "reverse" Stroop effect. Unfortunately, the authors do not fully address these issues in their discussion.

Association of Music Reading with Spatial Dimension

The Stroop effect is an asymmetric effect. Stroop (1935) demonstrated that irrelevant color names affected naming the ink color of the word, whereas irrelevant ink color did not affect word reading. A similar asymmetry was found also in the spatial version of the Stroop task (Lu & Proctor, 1995). In the spatial version of the Stroop task, participants are asked to respond to the spatial location of a written word that indicates a spatial direction (e.g., left) while ignoring the meaning of the word. The task presented by Grégoire et al. (2013) is similar to the reverse spatial Stroop task; namely, the relevant dimension is the word whereas the irrelevant dimension is its position on the musical staff.

Musical notation is a unique system that integrates all pitch dimensions (frequency, duration and volume) into single graphic symbols. Interestingly, music notation involves an additional dimension—spatial organization (location on the staff)—which is usually not associated with most other verbal reading systems. As far as we know this spatial aspect does not exist in other written systems. The codification of pitch frequency (which is in the center of interest in the present study) is based on the spatial position of the tone on the musical staff. Accordingly, high pitches are positioned on the higher part of the musical staff and low pitches are positioned on the lower positions of the staff. Following this view, the musical notation system can serve as a unique reading system where spatial positions are embedded in the graphical notation, providing important information about the musical pitch. In fact, the results presented by Grégoire et al. (2013) suggest that, in musicians, processing of spatial position is automatic and irrepressible even when it serves as an irrelevant dimension to the task performance.

The association between space and auditory tones was suggested by several studies. In 2006, Rusconi and her colleagues demonstrated the association between spatial response codes and auditory tones (Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). This was replicated a year later by Lidji and colleagues (Lidji, Kolinsky, Lochy, & Morais, 2007). These results suggest that musical (auditory) tones are mentally represented in a spatial array similar to the mental number line (Dehaene, 1997). The results presented by Grégoire et al. (2013) demonstrate that the association between pitch and spatial position is also inherent to the music written system. The codification of spatial position of musical notes on the staff is part of the music notational system and is automatically processed.

Interference vs. Facilitation

In their work, Grégoire and colleagues (2013) use note names that are printed outside a musical staff as neutral stimuli. These note names are compared with note names or words that are printed inside the musical staff, creating congruent or incongruent conditions. It is clear, as suggested by the authors themselves, that the neutral set of stimuli cannot serve as a baseline in order to assess the interference or facilitation components of the effect. However, Grégoire et al. do suggest that "the [congruency] effect is due to interference by assuming that the difference between the conditions is related to a perceptual complexity of the stimuli that is identical for musicians and non-musicians". According to the authors, “congruence elicits no beneficial effect in musicians whereas incongruence elicited detrimental effect”. As far as we understand, processing written notes on the staff is qualitatively different between these two groups. Namely, the irrelevant dimension (i.e., musical staff) has no meaning for non-musicians and for them the difference between *in context* and *out of context* stimuli could reflect an effect of perceptual complexity, as suggested by the authors. In contrast, for musicians, the two dimensions of the stimuli (spatial position on the staff and the word) are both inherent to the written music and to its reading. Accordingly, the difference between reaction times (RT) for the different conditions do not reflect the effect of perceptual complexity only, but also an effect that is related to the conflict between the information that is conveyed by the two dimensions of the stimulus. Therefore, the difference between the *out of context* and *in context* (i.e., existence of the staff) conditions is different between the two groups. It is quite possible that the congruency effect presented by the musicians is composed of both interference and facilitation, and the lack of a proper neutral condition prevents one from knowing what the contribution of each component to the total congruency effect is. Previous research showed that the Stroop effect is composed of both interference and facilitation (MacLeod, 1991). In the music domain, we have shown that auditory irrelevant tones can both interfere and facilitate responding to the written note in musicians (Akiva-Kabiri & Henik, 2012). It would be of interest to find out if the congruency effect found by Grégoire and colleagues is exclusively interference based, as the authors suggest, or whether it is also composed of a facilitation component. There is no a priori reason to assume that this effect should be totally dependent on interference. A possible future study will have to use an appropriate neutral condition (e.g., inverted staff) that can serve as a baseline in order to distinguish between facilitation and interference, at least for musicians.

Furthermore, it has been shown that in addition to the informational conflict in the Stroop task, there is another conflict—the task conflict (Goldfarb & Henik, 2007). The task conflict represents the conflict between the relevant and irrelevant tasks. In the Stroop task, the two competing tasks are color naming and word reading. In the work of Grégoire et al. (2013), it is possible that the task conflict is between reading the note word and naming the note according to its spatial position on the staff. This additional conflict is exclusive for the musicians since non-musicians are not able to name the musical tones. It has recently been suggested that facilitation is dependent on task conflict and is modulated by an individual's ability of control (Kalanthroff, Goldfarb, & Henik, 2012; Kalanthroff, Goldfarb, Usher, & Henik, 2013). Hence, examination of facilitation in the present context may be of general interest.

Using Musical Training as a Tool to Study Automatism

Another issue to consider in the paper by Grégoire et al. (2013) is the utility of musical training as a tool for studying the Stroop effect and automaticity. According to the authors, "musical formation provides much greater flexibility in manipulating practice than literacy learning. Control over practice is tighter...". However, considering the heterogeneity of music education and the relative little importance of verbal notation naming in the process of acquisition of musical expertise, we are not sure about the practical implications of such a manipulation on a rigorously controlled experiment.

It is not a secret that musical training is a heterogeneous process that could differ substantially from one musician to another. For example, musical education varies depending on the specific musical instrument that one is studying. Different musical instruments require different amounts of practice in order to master them. Also, the musical reading system could vary across instruments and disciplines; accordingly, it is possible to find different musical notation systems, such as tablatures, letters and numbers. Moreover, even within one notational system, such as the traditional Western notation, different instruments are often associated with different clefs (e.g., G, F and various C clefs), and therefore musicians studying various instruments may develop high proficiency in reading one clef and not another.

Even when control is tight, musical note naming per se is not a fundamental activity in musical training. In certain disciplines (i.e., Suzuki System) playing an instrument is acquired by playing by ear before the acquisition of music reading. In general, when taking music lessons, one is studying to associate the different dimensions of the musical tone with a certain body or hand movement in order to produce a coherent auditory output. In addition, unlike other reading systems, music reading can involve several different alternatives of transcoding of the visual symbol (i.e., verbal, motor, or auditory outputs; Schön, Anton, Roth, & Besson, 2002). Verbal output of music reading (i.e., labeling the musical note) is sometimes only a marginal ability in musical training. Hodges (1992) suggest that even if it is clear that music reading is an invaluable ability in expert musicians, it is not clear to what extent music reading is considered to be a basic part of music education for many music teachers. Hence, the type and amount of practice with musical note reading is not the main activity in musical training and seems impractical to control, since musical training is highly heterogeneous.

Taken together, it seems that musical note reading does not necessarily reflect the main activity of music making and varies substantially between disciplines of music teaching and instruments, especially during the initial phases of music acquisition but also at later phases. Since musical training is highly heterogeneous, it seems impractical to control all the possible confounds that are related to musical education and practice. Hence, it is not clear what research advantage musical training has over training of any second language reading, where the learning process is more homogeneous and the association between automaticity and training can be studied at various chronological ages and levels of expertise.

Conclusions

In sum, in their paper Grégoire et al. (2013) demonstrate the effect of irrelevant spatial position of musical notes on reading note word names by musicians. It is interesting since music notation is a unique system that integrates spatial position with other graphical symbols. Music processing and musical training are fascinating processes that deserve to be studied. However, the contribution of musical training in order to study the acquisition of automaticity is not clear and may be impractical due to the high heterogeneity of music training. For the purpose of studying automatisms, it might be easier to examine changes in automaticity in second language acquisition.

References

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology*, *59*(5), 272-278.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* New York: Oxford University Press.

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1170-1176.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*(4), 269–278.

Hodges, D. A. (1992).The acquisition of music reading skills. In R. Colwell (Ed.), *Handbook of research in music teaching and learning* (pp. 466-471). New York: Schirmer Books.

Kalanthroff, E., Goldfarb, L., & Henik, A. (2012). Evidence for Interaction between the Stop Signal and the Stroop task conflict. *Journal of Experimental Psychology: Human Perception and Performance*. E-pub ahead of print.

Kalanthroff, E., Goldfarb, L., Usher, M., & Henik, A. (2013). Stop interfering: Stroop task conflict independence from informational conflict and interference. *Quarterly Journal of Experimental Psychology.* E-pub ahead of print.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1189-1207.

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*(2), 174-207.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, *109*(2), 163-203.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113-129.

Schön, D., Anton, J. L., Roth, M., & Besson, M. (2002). An fMRI study of music sightreading. *NeuroReport*, *13*(17), 2285-2289.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

Annexe 2

What is Learned, and When? Commentary on “The Musical Stroop Effect: Opening a new Avenue to Research on Automatisms" by Grégoire, Perruchet, and Poulin-Charronnat (2013)

Anne Gast

Abstract

In the musical Stroop task, which has recently been introduced by Grégoire, Perruchet, and Poulin-Charronnat (2013), participants respond to note names that are placed inside musical notes. Musicians respond more slowly to note names that are incongruent with the note than to note names that are congruent with the note. Grégoire et al. propose to use this task to study the acquisition of automaticity by relating musical Stroop effects to the amount of musical experience. I discuss some caveats that have to be considered for these types of analyses. Specifically, I focus on how different contingencies in the learning situation relate to the Stroop effect and on the question whether a long-term perspective is suitable for studying the acquisition of automaticity.

In a recent paper, Grégoire, Perruchet, and Poulin-Charronat (2013) introduced a new Stroop-like task, the "musical Stroop" paradigm. In this task, names of musical notes ("Do", "Re", "Mi", etc.) appear written inside musical notes that are positioned on a musical staff. Depending on the position of the note on the staff, note name and note are congruent or incongruent with each other. Participants either have the task to indicate whether the note name refers to a note or to read out the note name. With both tasks, a musical Stroop effect was shown for participants who had several years of musical training: They responded more slowly to note names that were incongruent to the note's position on the staff than to note names that were congruent to the note's position on the staff.

In the musical Stroop task, the note has the role of the task-irrelevant and interfering stimulus feature (which, in the classical Stroop task, is the meaning of the color word). Although many versions of the Stroop task have been published, only relatively few have been successful at replacing the task-irrelevant (and interfering) stimulus feature (see MacLeod, 1991, for a review). The musical Stroop effect is therefore an important generalization of Stroop effects.

According to the authors, the most important advantage of the musical Stroop task, however, is that it offers new possibilities to study the acquisition of automatisms by relating it to years of musical experience. With this, the authors direct attention to two important issues: automaticity is most likely gradual and it is usually based on learning. Repeated practice with a task leads to automatization of the practiced behavior (e.g., Logan, 1988; McLeod, 1991; Moors & De Houwer, 2006; Schneider & Shiffrin, 1977). Studying the development of a musical Stroop effect in relation to musical experience, could therefore yield important insights into automaticity in general. The resulting learning curve could be informative about the general speed of acquisition of automaticity, about changes in acquisition speed (e.g., whether and when the Stroop effect reaches an asymptote), and about what happens if people stop or re-start musical practice. Furthermore, it is interesting to relate acquisition of note-reading automaticity to the high-level skill that it serves – mastery of a musical instrument. The authors argue that the musical Stroop effect is more suitable for these types of analyses than classical Stroop effects because musical education–in contrast to learning to read–starts at any age and is therefore easier to decouple from confounding variables, such as age or cognitive development.

The present commentary focuses on this promise of the musical Stroop task. A tool that allows investigating the effect of long-term practice on automaticity without expensive and time-consuming training sessions in the lab, could yield many important new insights into the nature of automaticity in general. A few points need to be considered, however, if the musical Stroop task is to be used in this way. I will focus on two aspects: First, what is actually learned in musical education and how does this relate to what is picked up by the musical Stroop task? Second, are the proposed macro-level analyses suitable given the typical time course in which automaticity is acquired?

What is learned?

Grégoire et al. (2013) refer to *note naming* as both the behavior practiced during musical education and as the interfering response in the musical Stroop task. It is helpful, however, to analyze in detail the different contingencies that might be encountered in the learning situation. First, students might learn to give the correct response upon seeing a note; depending on the context, this could be to name the note or to produce the correct tone (pitch) on an instrument. Learning these responses could be called stimulus-response (S-R) learning. Second, musical students also learn relations at the stimulus level, for example they acquire knowledge on how notes are called (the relation of note and note name), and to which tones the note and the note name refer to. This could be called stimulus-stimulus (S-S) learning.

In the musical Stroop task with the note name reading task, one type of previous S-R-learning as well as previous S-S-learning can contribute to the effect. This is the case because in the musical Stroop task with note name reading, two types of compatibility co-vary between congruent and incongruent trials: First, the S-R compatibility of the note name and the naming response that has been learned to give upon presentation of the note (note response) and, second, the S-S compatibility of note name and note. As is the case in the classical Stroop task, S-R and S-S compatibility are therefore confounded in the musical Stroop with the note name reading task (see De Houwer, 2003a, b). While S-R-compatibility effects are usually explained by an interfering or facilitating effect on response selection, S-S-compatibility effects are usually attributed to the stimulus encoding phase.

Grégoire et al. (2013), however, also show the musical Stroop effect with a go/no-go task in which participants have to press a key if a word is a note word. Crucially, effects from this task can only be based on S-S compatibility; S-R compatibility cannot contribute to the effect. This is the case because both in congruent and in incongruent trials, the note name (and also the note) obviously always refers to a note and thus always to the go-response (no-go trials exist, but are, of course, irrelevant for the calculation of the Stroop effect).

It is interesting to note that musical Stroop effects from the go/no-go paradigm (which can only be based on S-S-compatibility) were larger than effects from the note name reading paradigm. This suggests that the reported effects are largely due to learned S-S compatibility and not to learned S-R-compatibility. What primarily drives the effect seems not to be the automaticity of note naming (at least if this refers to the automaticity of saying the note name when seeing a note), but the learned compatibility of note and note name. Presumably, for musicians, incongruent notes interfere with the encoding of note words.

This does not make the effect uninteresting. To the contrary, it can be argued that the effect is more process pure and can therefore tell us more about the learning situation from which it resulted than the classical Stroop effect. It also does not indicate that musical training is generally based on S-S learning only. Rather, the proposed task versions are not ideally designed to pick-up another probably highly automatized S-R-link after years of musical training: Playing the correct tone on an instrument when seeing a note. It is possible to design a musical Stroop task that can pick up this relationship. Participants, instead of reading note names, could respond to note names by playing on their instrument one of the tones to which the note name refers. If presented along with incongruent notes, the strongly practiced playing-response might interfere and lead to a potentially larger and more reliable Stroop effect, which might even be more suited for the differential analyses proposed by the authors.

It is helpful to keep in mind which (S-S or S-R) contingencies could have been learned and which of these can be picked up by a specific Stroop task. In general, different relations could be acquired in different pace. If their impact is not discriminated, potentially interesting steps of automatization might remain undected. Furthermore, acquiring S-S vs. S-R relations in musical training is probably based on different types of learning. S-S learning might require only the presentation of the two stimuli. Learning to respond to a note, however, begins with intentionally generating the response and thus has the structure of operant conditioning, which requires a reinforcer. Using different versions of the musical Stroop task might be helpful in distinguishing the impact of different learned relations.

When is automaticity learned?

The feasibility of studying the acquisition of musical automaticity by relating it to musical experience also depends on the time course in which musical automaticity is acquired. First, can we assume that the musical training proceeds in a steady fashion? As argued above, musical training does not consist of the uniform training of note reading.

Different phases of the training might focus on different aspects. Especially if typical stages of musical education exist across students, it is difficult to draw conclusions about the long-term acquisition of automatisms. For example, if it was found that the musical Stroop effect increases during the first year of musical exercise and then reaches an asymptote, can we conclude that this is informative about acquisition of automatisms in general, or could this be due to a change in what is typically practiced in the first vs. second year of musical training?

A second issue concerns the time frame in which the theoretically interesting changes in automaticity take place. Undoubtedly, the amounts of practice from years of musical training are difficult to achieve in the lab. These large amounts of training, however, cannot be quantified precisely because they do not only depend on the exact duration of musical education, but also on frequency of practice, etc. Assessing existing differences in musical experiences thus zooms out to a macro-level of practice but lacks micro-level information that would be available from experimental training sessions (e.g., number of practice trials). Is the macro-level adequate for reaching an understanding of the acquisition of automaticity?

A few studies showed that several training sessions with new stimulus-response contingencies can produce Stroop-like effects (see MacLeod, 1991). For example, MacLeod and Dunbar (1988) showed that newly learned naming responses can lead to interference in a Stroop task after five one-hour training sessions. Other attempts at training an interfering stimulus dimension have been less successful with comparable amounts of training (e.g., MacLeod, 1998).

Even faster acquisition of interfering automatisms has been shown in the literature on cognitive control. In this area, typically two tasks with arbitrary stimulus-response mappings are combined with each other in a dual-task situation. Response times on stimuli for which the other task would have specified the opposite response are typically slower. These effects are structurally similar to S-R compatibility effects in the Stroop task. As stimulus-response mappings are arbitrary, however, they cannot be based on long-learned automatisms, but must have been acquired within a single experimental session (see Meiran, Cole, & Braver, 2012).

Information on the acquisition of automaticity is also available from paradigms that assess automaticity in a different manner. In studies by Logan and colleagues (e.g., Logan & Klapp, 1991), the criterion of automaticity is not interference with another task, but whether response latency ceases to depend in a linear fashion on the number of necessary algorithmic steps. Logan and Klapp showed that, given appropriate training, automaticity can be reached within a single training session of 15 minutes.

As the reported musical Stroop effects are probably largely due to S-S-compatibility, it is interesting to consider learning effects that are largely based on S-S learning, such as some types of Pavlovian conditioning. In evaluative conditioning, for example, there is strong evidence that S-S learning can lead to automatic behavioral effects (on an implicit measure) within a few trials or even only due to instructions (e.g., De Houwer, 2006).

To summarize, across research areas there is evidence that some aspects of automaticity can be learned relatively quickly. Although the necessary amount of training differs widely depending on conditions that are not yet well understood, this suggests that important steps of automatization acquisition might be missed out when comparing people with different amounts of musical experience.

Conclusions

Grégoire et al. propose an interesting new venue for studying the acquisition of automaticity by relating musical experience to the size of a musical Stroop effect. This approach not only avoids time-consuming training sessions, it also might reveal aspects of automaticity that only show up after amounts of training that are difficult to reach in the lab. I discussed two points that should be considered when applying this approach.

First, while Grégoire et al. refer to the automaticity of note naming, there are several S-S and S-R relations that are learned in musical education and that could contribute to the musical Stroop effect; note-naming might not be the most relevant one. This complicates the interpretation of which automatized response or process underlies a given Stroop effect. Varying the type of Stroop task, as done in the target article, however, can give important information on what exactly was automatized in a complex learning situation.

Second, the number of years of musical experience gives a long-term perspective, but is also only a rough estimate of the actual amount of musical training. Theoretically interesting steps of automatization often take place within few trials, and might thus be missed out with these analyses. More research is necessary to predict when a long-term analysis of experience is promising and when a more fine-grained perspective is necessary. Such a perspective could be reached by following musical students closely during their first weeks of practice or by experimental training sessions that allow trial-to-trial analyses.

References

De Houwer, J. (2003a). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The Psychology of Evaluation: Affective Processes in Cognition and Emotion* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J. (2003b). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition*, *31*, 353-359.

De Houwer, J. (2006). Using the implicit association test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation, 37*, 176-187.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527. doi: 10.1037//0033-295X.95.4.492

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: 1. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 177–193.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, *26*, 201-211.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,*126-135.

Meiran, N., Cole, M. W., & Braver, T. S., (2012). When planning results in loss of control: intention-based reflexivity and working-memory. *Frontiers in Human Neuroscience*, *6*, article no: 104.

Moors, A. and De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297-326.

Schneider, W, & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,*1-66.

Annexe 3

How automatic is the Musical Stroop Effect?
Commentary on *The Musical Stroop Effect: Opening a new Avenue to Research on Automatisms?* by Grégoire, Perruchet, and Poulin-Charronnant (2013)

Birte Moeller & Christian Frings

Abstract

Grégoire, Perruchet, and Poulin-Charronnant (2013) investigated a musical variant of the reversed Stroop effect. According to the authors, one big advantage of this variant is that the automaticity of note naming can be better controlled than in other Stroop variants as musicians are very practiced in note reading whereas non-musicians are not. In this comment we argue that at present the exact impact of automaticity in this Stroop variant remains somewhat unclear for at least three reasons, namely due to the type of information that is automatically retrieved when notes are encountered, due to the possible influence of object-based attention, and finally due to the fact that the exact influence of expertise on interference cannot be pinpointed with an extreme group design.

The Stroop effect, first reported by Stroop (1935), includes the observation that the naming of the ink color of color-words is influenced by the congruity between word's meaning and the word's ink color. In the typical Stroop task (for a review, see MacLeod, 1991), participants report the ink colors of words whose meaning is either congruent or incongruent to the ink color. The naming of ink colors is fast and participants make few errors if ink color and word meaning are congruent and it is slow and participants make more errors if ink color and word meaning are incongruent (see e.g., Dunbar & MacLeod, 1984). A prominent account as to explain the Stroop effect is an automaticity approach, that is, it is assumed that both dimensions of a Stroop stimulus (ink color and meaning of the word) are processed but that they require different amounts of cognitive resources. Word reading (due to practice) is a nearly automatic process whereas naming of ink colors is not. More automatic processing then interferes with less automatic processing, while less automatic processing interferes only to a smaller degree with more automated processing. In fact, color-word reading can also be hampered by an incongruent ink color of the presented word (e.g., Dunbar & MacLeod, 1984), a phenomenon labeled the reverse Stroop effect, that is typically not as large as the classic Stroop effect.

Grégoire, Perruchet, and Poulin-Charronnant (2013) investigated a variant of the reverse Stroop effect. They found that word reading (names of notes) was hampered by pictures of incompatible notes for musicians who frequently name notes, but not for non-musicians who have no experience with musical notation. This variant is very similar to he picture-word Stroop variant (participants typically have to name a picture while a congruent or incongruent word is presented on the picture; reverse Stroop effects can be observed with this variant as well, e.g., Glaser & Düngelhoff, 1984). Grégoire et al. suggested that the musical variant of the Stroop effect is especially interesting as the degree of automaticity can more easily be controlled than with other stimulus materials because the irrelevant dimension (the musical notation) is completely unpracticed for non-musicians whereas it is highly practiced for musicians. In this comment, we discuss three aspects that possibly undermine the arguments of Grégoire et al. (2013) concerning automaticity.

Automatic retrieval of motoric sequences rather than automatic note naming

First, a difficulty with the suggested paradigm is to determine which automatic responses are induced by the printed notes. In contrast to the stimulus materials in the standard Stroop task, an exceptional property of notes is that they refer to motoric sequences other than pronunciation of words. Depending on the instrument of a musician, a note indicates a certain motoric pattern (e.g., a sequence of finger, arm or lip movements) that is necessary to produce the tone depicted by the note. In fact, on the great majority of encounters of a printed note, a musician will respond by producing the corresponding tone on his or her instrument. Importantly, in order to produce a tone, indicated by a note, naming of this note is not necessary (and in some cases not even possible - just think of a trumpeter!). Therefore, for most musicians, it seems likely that the motoric response that is required to produce the tone on the musician's instrument and not the pronunciation of the note name is overlearned and therefore automatically retrieved at the encounter of a printed note. In addition, depending on the particular instrument, very dissimilar motoric responses will be associated with a particular note (e.g., consider the differences between violinists, trumpeters, and pianists). But even concerning one instrument, say a violin, the same note can be produced by different motoric patterns, for example using different fingers and/or different strings to produce the same tone. From this perspective, it is difficult to determine how automaticity influenced interference in the musical Stroop effect as analyzed by Grégoire and colleagues. In fact, interference between a note and a written note name seems to take place on a semantic level whereas the automatic response retrieved by the note is foremost motoric and possibly ambiguous. With a sample of musicians who play various instruments, it is clear that interference occurred at a semantic level while it remains unclear how the different motoric patterns automatically retrieved by the musicians influenced this interference. One way to determine this influence would be to change the task, allowing interference to take place at the motoric level. In fact, such interference has been shown by Zakay and Glicksohn (1985), who first reported a Stroop like effect with musical material, using a sample of pianists and finding an influence of automatic processes on key presses on the piano (for another study showing a reverse Stroop effect with musical material, see Akiva-Kabiri & Henik, 2012).

Modulation due to object-based attention

Second, Grégoire and colleagues (2013) suggest that their effects are due to an automaticity of note processing in musicians. Yet, it can be argued that such an automatic processing of the notes might be specific to the particular design of the stimuli used. Wühr and Frings (2008; see also Wühr & Waszak, 2003) found that 'automatic processing' in a Stroop like task may be modulated by object-based attention, rather than being automatic. The authors presented colored objects on a colored background and asked the participants to name the color of a relevant object. Additional color-words were presented either in the relevant or the irrelevant object or in the background. Stroop interference by the color-word was strongest if the word appeared in the target object and weakest if the word appeared in the irrelevant object with intermediate interference by color-words in the background. Other variants of the Stroop effect, for example the Emotional Stroop effect, were confined to a condition in which the irrelevant and relevant dimensions were both parts of the same object (cf. Frings & Wühr, 2012). Similar to the one-object condition in the study by Wühr and Frings (2008), the note and note name presented by Grégoire and colleagues always belonged to the same object. Thus, it cannot be ruled out that the interference of note naming on note name reading would diminish if written note name and note would not belong to the same object. This seems to be the case for other variants of the Stroop effect and would argue against a strong claim of automatic processing. To determine whether object-based attention modulates automatic interference of note name reading due to additional notes, an experiment might vary whether the name is or is not printed as a part of the note. **Figure 1** depicts the basic idea of presenting note and note name as part of the same or different objects.

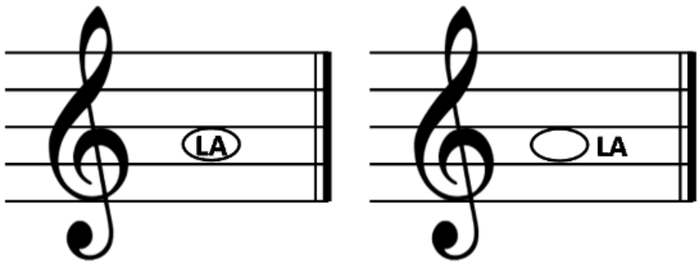

Figure 1. A sketch for possible stimuli that may be used to investigate the influence of object based attention on automatic note naming: name and note are presented as one object in the left, as in the experiments of Gregoire et al. (2013), and as two separate objects in the right illustration. Obviously, in this sketch spatial attention is confounded with object-based attention.

Influence of expertise on the musical Stroop effect

Third, one argument of Grégoire and colleagues (2013) for using musical notation is that automaticity of note naming is relatively easy to determine via the degree of expertise. In this regard, the obvious question is, what kind of influence we can expect practice to have on the musical Stroop effect as defined by Grégoire and colleagues. On the one hand one might expect more interference with more automaticity of note naming (i.e., with more expertise). On the other hand, in the picture-word interference task, which in our view is a close neighbor of the musical Stroop effect, low-frequency words have been shown to interfere more with picture naming than high-frequency words (Dhooge & Hartsuiker, 2010; Miozzo &

Caramazza, 2003; for the same effect in color naming see Burt, 2002; but see Klein, 1964). Thus, assuming that expertise includes being frequently confronted with a certain kind of items, one might speculate that no expertise results in no interference, lower levels of expertise might lead to interference, and higher levels of expertise might result in less interference, leading to a U-shaped relation between expertise and the amount of interference. In the study by Grégoire and colleagues the amount of practice varied between a considerable amount of practice in one group and no familiarity at all with musical notation in the other. For participants who do not know the meaning of printed notes it was reasonable to expect that no interference on note name reading could take place, which is exactly what the authors found. Similarly, the authors expected and found interference for participants who have years of practice in note naming. However, investigating only these two extreme levels of expertise it cannot be determined whether interference monotonously increases or rather follows an inverted U-shaped function with more practice with the notation. As Grégoire and colleagues point out, musical notation can be varied quantitatively by using additional samples with different degrees of expertise. However, since the authors did not include another sample, the exact influence of automatic note naming by musicians on word reading cannot be pinpointed at this point in time.

Conclusion

In summary, Grégoire and colleagues (2013) report a possibly interesting variation of a picture-word interference task which emphasizes the impact expertise might have on Stroop like interference. While the task in its present format raises some doubts concerning the role of automaticity in producing the observed interference, it also might open interesting avenues for research on (motoric) interference in Stroop-like tasks in the future.

References

Akiva-Kabiri, L. & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology, 59*, 272-278.

Burt, J. (2002). Why do non-color words interfere with color naming? *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1019-1038.

Dhooge, E., & Hartsuiker, R. (2010). The distractor frequency effect in picture-word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 878-891.

Dunbar, K., & MacLeod, C. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human, Perception and Performance, 10*, 622-639.

Frings, C., & Wühr, P. (2012). Don't be afraid of irrelevant words: The emotional Stroop effect is confined to attended words. *Cognition and Emotion, 26*, 1056-1068.

Glaser, W. & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640-654.

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Klein, G. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77*, 576-588.

MacLeod, C. (1991). Half a century of research on the Stroop tasks: The progression of interference and facilitation. *Psychological Bulletin, 109*, 163-203.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General, 132*, 228-252.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Wühr, P., & Frings, C. (2008). A case for inhibition: Visual attention suppresses the processing of irrelevant objects. *Journal of Experimental Psychology: General, 137*, 116-130.

Wühr, P., & Waszak, F. (2003). Object-based attentional selection can modulate the Stroop effect. *Memory & Cognition, 31*, 983-994.

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology, 39*, 414-423.

Annexe 4

Can the "Musical Stroop" task replace the classical Stroop task?
Comment on: Gregoire, Perruchent and Poulin-Charronnat (2013) "The Musical Stroop Effect: Opening a New Avenue to Research on Automatisms"

Dan Zakay

Abstract

The musical Stroop task is analyzed and compared to the classical Stroop task. The analysis indicates that the two tasks differ in the following significant characteristics: ecological validity, the interrelations between the two perceptual dimensions involved, the nature of the automatic process and the existence of a potential Garner interference. It is concluded that the musical task has no advantage over the classical task.

In 1935, Stroop designed the color-ward task (Stroop, 1935), which turned to be, later on, one of the most used paradigms in cognitive research. Since then, many researchers have tried to develop new tasks, hoping that it can equal the power of the classical Stroop task (CST). However, all of the new tasks lack some of the critical features (as detailed below) of the CST, and because of this such tasks are labeled STROOP-LIKE tasks (for a review see Macleod, 1991). Zakay and Glickshon (1985) introduced another Stroop-like task: the musical-Stroop-task (MST), in which the stimuli are either notes or names of notes printed in congruent or non-congruent positions on the staff. Participants were required to respond verbally by reading the name of the note or manually by pressing appropriate piano keys. Gregoire, Perruchent and Poulin-Charronnat (in press) utilized the MST in order to find a Stroop effect, which is equivalent to the effect usually obtained in the CST, in the musical domain. They did so by introducing some changes to the original MST, and by adding new experimental control conditions to the procedure. In the revised MST, participants responded only verbally, and a new kind of stimuli was added: a note name written within a note picture.

The authors suggest that the revised MST is opening a new avenue to research on automatisms (title of the paper). Furthermore, they argue that the revised MST is even better, in this respect, than the CST.

In this comment I negate this ambitious conclusion and indicate why the revised MST is no more than another Stroop-Like test. I am arguing that the two changes made are decreasing the power of MST as a test of automaticity and its degree of similarity with the CST. These arguments are based on an analysis showing that the revised MST is essentially different from the CST in several important aspects.

Ecological validity

Whereas reading of words is a natural behavior of most people, reading names of notes appearing on a staff is an unusual behavior, even for musicians. The natural behavior of musicians when notes are presented to them is the production of music by operating designated controls (e.g. Keys) of musical instruments, and this includes singing (see Palmer, 1989; 1997). Pictures of notes are not symbols which represent words but rather symbols which trigger certain psycho-motor procedures by which musical instruments are operated. When names of notes are written within notes pictures on a staff, a very peculiar and unusual non-ecological condition is created. Musicians are never facing such stimuli and task. A manual response on an appropriate piano key, as was done by Zakay and Glickshon (1985), is a much better response than reading, from an ecological-validity perspective.

The interrelations between the two perceptual dimensions

In the CRT, the two perceptual dimensions are embedded within each other in the strongest possible way. This is creating an intensive conflict between the outcomes of the two processes involved in the processing of the two dimensions. The outcomes are of the same kind, namely, a name of a color. This is not the case in the MST since the two dimensions are a ward and a location of the word, or of a note, on the staff. The location is an external dimension of the word or the note and is not an integral part of it, Like in the case of the color of an ink in which a color name is written. Furthermore, the outcomes of the processing of the note name or the note and of the location are not necessarily the same. The outcome of processing the location, as explained earlier, is a tendency to act and activate a psychomotor action. Thus, the intensity of the conflict between the word and its location is not high.

The nature of the automatic process

Whereas in the CST it is clear that reading words is an automatic process (e.g., the word superiority effect, Raap, Newsome, McDonald & Scvaneveldt, 1982), this is not the case in the MST. As explained earlier, when a musician sees a note he/she tends to start a motor process, not a reading process. Thus the automatic process is that of activating the motor process, but this is not part of the revised MST. Thus, if one wants to test automaticity, the manual operation used by Zakay and Glickshon (1985) is a much more adequate response than reading.

A potential Garner interference

One danger to the interpretation of Stroop effects is that the effects are reflecting basic differences between the perceptual dimensions involved, in terms of the ease of perceptual processes of each dimension by itself (A Garner interference). Pansky and Algom (1999) demonstrated that unless two perceptual dimensions are free of Garner interference among themselves, the meaning of a Stroop effect might be problematic. This is a challenge for the MST because it should be demonstrated that the processing of notes pictures, notes names and locations on a staff, each one by itself, are equal in perceptual processing difficulty.

Conclusions

The revised MST is an interesting replication and development of the Zakay and Glickshon (1985) task. There might be a potential use of the MST in order to analyze varing degrees of automaticity de-confounded from chronological age. This potential use should await validation and is highly dependent on the allocation of appropriate populations. Nevertheless, the MST should be considered a Stroop-like task, and is certainly not a substitute of the classical Stroop test.

References

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology, 60*, 269–278.

Macleod, C.M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*, 163-203.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15*(12), 331-346.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48*, 114-138.

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 39-59.

Raap, K.R., Newsome, S.L., McDonald, J.E., and Scvaneveldt, R.E. (1982). An activation-verification model for letter and word recognition: the word-superiority effect. *Psychological Review, 89*(5), 573-594.

Stroop, J.R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology, 18,* 643-662.

Zakay, D., and Glickshon, J. (1985). Stimulus congruity and s-r compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology, 39*(3), 414-423.