EFFECTS OF AGE AND EDUCATION ON THE STROOP INTERFERENCE[1]

Raymond Bruyer
Université de Louvain
Martial Van der Linden
Université de Liège
Dominique Rectem and Clelia Galvez
Cliniques universitaires Saint-Luc, Bruxelles

Typically, the naming of the colors in which words are printed is slower when the words are conflicting color names (e.g. the word «red» printed in green, as compared to the same word printed in red or in black): this phenomenon is known as the «Stroop effect» or «Stroop interference» (after Stroop, 1935). In a recent impressive review of over 400 publications on the Stroop effect, MacLeod (1991) detected «eighteen major empirical results that must be explained by any successful account on the Stroop effect» (listed in MacLeod, 1991, Appendix B, p. 203). Among these results, the author mentioned the existence of changes in the Stroop interference effect due to normal aging. Indeed, several studies have shown that the elderly (i.e., over 60 years old) tend to display larger Stroop interference effects than younger subjects (e.g., Cohn, Dustman & Bradford, 1984; Comalli, Wapner & Werner, 1962; Dulaney & Rogers, 1994; Panek, Rush & Slade, 1984).

A recent paper by Houx, Jolles and Vreeling (1993) verifies this detrimental effect of age on the Stroop interference. In addition, the authors also showed that this increase of interference with age is especially observed in subjects who had suffered «biological life events» (i.e. mild biological or environmental factors, such as undergoing repeated general anesthesia, susceptible of altering brain function). Moreover, Houx *et al.* (1993) reported an effect of education, since the effects of age and of age x presence/absence of biological life events were amplified in subjects with a low level of education, and a direct measure of interference (see below) showed a significant main effect of the education level. Thus, it could be that, in a near future, a new «major empirical result» will be added to the series of eighteen empirical data that must be accounted for by any theoretical model of Stroop interference (MacLeod, 1991).

[1] This study was made in part thanks to the grant 92/97-158 («Action de Recherche Concertée») of the government of the French-language community of Belgium.

More generally, these studies contribute to the opinion according to which many age-associated cognitive deficits would be better understood by taking different variables into account, especially subject variables such as education, health status, intellectual activities, and occupation (among others). For example, a study by Arbuckle, Gold, Andres, Schwartzman and Chaikelson (1992) showed that being younger, healthier, more educated, more introverted, more intellectually active and more satisfied with social support predicted less intellectual decline and better memory performance (see also Arbuckle, Gold & Andres, 1986; Craik, Byrd & Swanson, 1987; Van der Linden, Wyns, Bruyer, Ansay & Seron, 1993). Nevertheless, other studies (reviewed in Salthouse, 1991) indicated that age effects on cognitive functions are relatively independent of background variables such as health, intellectual activities, occupational status or education.

With regard to education, Salthouse (1991) concluded: «at the present time, therefore, it does not appear that variation in education can account for more than a small proportion of the age differences observed in certain measures of cognitive functioning» (p. 77). Given these opposing theses, and since only one (to our knowledge) published research paper investigated education effects on the Stroop interference (Houx *et al.*, 1993), the main purpose of the present study was to re-examine the effects of age and education level on the Stroop test. An additional objective was to compare different measures of the Stroop interference. Indeed, as far back as 1965, Jensen detected in the literature no less than eleven different types of scoring the Stroop test, and other indices were proposed later (for instance Houx *et al.*, 1993 and the present study).

Method

Subjects

Ninety-nine normal subjects (60 females, 39 males) were enrolled. The group of young subjects consisted of 52 subjects (mean age = 24.75; range: 18-35) and the group of elderly subjects was formed with 47 subjects (mean age = 68.85; range: 60-79). Subjects of both age groups were classed according to three education levels, defined as follows: low level refers to schooling limited to primary (max. age: 12 years) or professionnal education; middle level relates to secondary school education (max. age: 18 years); high level concerns college and university education. Table I displays the demographic properties of the resulting six subgroups.

Within each age group, the three subgroups (education level) were compared with regard to their age (one-factor analysis of variance: ANOVA). The test did not reach a statistically significant level for young ($F(2,51) = 1.72$, NS) and for elderly subjects ($F(2,46) = 0.021$, NS). The same three subgroups were also compared with regard to the distribution of subjects in two gender categories (chi^2 tests). No test proved to be significant

Table 1: Subjects.

Age group	Sociocultural level	Number of females	Number of males	Mean age	Age range
young	low	8	5	23.62	18-35
	middle	6	4	27.00	21-33
	high	20	9	24.48	20-33
elderly	low	8	8	69.12	60-79
	middle	7	7	68.79	60-79
	high	11	6	68.65	60-79

(low level: $chi^2 = 0.386$, df = 1, NS; middle level: $chi^2 = 0.235$, df = 1, NS; high level: $chi^2 = 0.088$, df = 1, NS). In addition, within each education level, the two age groups were compared with regard to the distribution of subjects in two gender categories. The test was not significant, in young ($chi^2 = 0.377$, df = 2, NS) as well as in elderly subjects ($chi^2 = 0.949$, df = 2, NS).

Stimuli and procedure

Each subject was submitted, in a single session, to a close adaptation of the test initially devised by Stroop (1935). The session consisted of three stages, in the following order[2]. First, a series of 100 colored squares (randomly red, green or blue) was shown and the subject had to name the colors as quickly as possible without error (condition «N» for «naming»). Second, a series of 100 color names (randomly «red», «green» or «blue», written in black; actually, the french words *rouge*, *vert* and *bleu* were used) was shown and the subject had to read them aloud as quickly as possible without error (condition «R» for «reading»). Third and finally, a series of 100 colored color names was displayed and the subject had to name the color of the words as quickly as possible and without error, irrespective of the meaning of the word (condition «I» for «interference»). The names «red» (*rouge*), «green» (*vert*) or «blue» (*bleu*) were colored in red, green or blue and dis-

2 Actually, Stroop (1935) used the three conditions in two separate experiments, submitted to independent samples of subjects. Indeed, condition R was involved in Exp. 1 (with another condition not used here) while conditions N & I were used in Exp. 2. In addition, the order of conditions N & I was counterbalanced across participants. In the present study, however, the three conditions were administered in a fixed order, in a within-subject design. We agree that a manipulation of the order of the conditions was a better strategy, but this procedure results from the fact that the «Stroop test», actually, is part of a larger individual neuropsychological screening.

played at random, with one constraint: there was never any congruency between the meaning of a word and its color.

In each condition, the response time (in sec) was recorded[3] as well as the errors (with *vs* without correction by the subject).

Results

Table II shows the mean results for each subgroup of subjects. Even if statistical analysis of errors does display complex effects and interactions, caution is required because the accuracy was generally very high in every subgroup and condition (general mean errors = 0.734 out of 100; range = 0-3.385). As a matter of fact, the subjects performed very accurately so that the main dependent variable will be the speed to perform the tasks (hereafter «response time»).

Table II: Mean results for each subgroup in the three experimental conditions.

	Young			Elderly		
level	low	middle	high	low	middle	high
Time (sec)						
color naming	60.92	60.10	60.76	73.87	73.00	63.00
reading	48.08	38.30	41.38	48.00	48.50	45.35
interference	111.23	100.70	101.52	147.00	156.00	113.47
Corrected errors (/100)						
color naming	0.69	0.70	0.55	1.12	0.93	0.82
reading	0.23	0.10	0.21	0.00	0.00	0.06
interference	2.38	1.50	1.31	2.50	1.36	1.47
Non-corrected errors						
color naming	0.23	0.00	0.28	0.31	0.14	0.00
reading	0.15	0.00	0.00	0.00	0.07	0.00
interference	3.38	0.00	1.03	0.00	0.21	0.06

Accuracy

The number of errors (Table II) was studied by means of a four-way ANOVA. The two between-subject factors were age and education, and the

[3] Thus, this dependent variable was the total time necessary to respond to the entire sequence of 100 stimuli, not the RT of individual responses. The rationale of this procedure was twofold: (a) the subject saw the entire sheet with 100 stimuli, and (b) it was the procedure originally used by Stroop (1935).

two within-subject factors were condition and kind of error (corrected *vs* non-corrected). Only significant effects will be reported ($p = 0.05$ or less) and *post-hoc* analyses were computed by means of the Newman-Keuls test ($p = 0.05$ or less). The number of corrected errors was higher than the number of non-corrected errors (0.886 *vs* 0.583: $F(1,93) = 29.81$, $p < 0.0001$), with low level subjects producing a greater number of errors (0.918) than high (0.483) and middle level subjects (0.417) not differing from each other ($F(2,93) = 4.61$, $p < 0.02$), and errors were less numerous in condition R (0.48) than in the other two conditions (N: 0.783; I: 0.94) not differing from each other ($F(2,186) = 29.59$, $p < 0.0001$). Four significant interactions involving two factors, two involving three factors and the interaction involving the four factors also reached a significant threshold. However, given the general low number of errors, these interactions were not analyzed any further. We only mention, as can be seen in Table II, that the highest number of errors was observed for non-corrected errors in condition I submitted to low level young subjects; more generally, the highest numbers of errors were observed in condition I for low level subjects.

Response time

The response time in the three experimental conditions (Table II) was studied by means of an age x education x condition ANOVA. The dependant variable was the natural logarithm transformation of the response time (but for clarity, the means reported in what follows will be expressed in sec). This transformation was used to stabilize variances and to normalize distributions, but each following analysis was computed on both the transformed and the untransformed times. In what follows, F_T concerns the untransformed time and F_L the logarithmic transformation of the response time. *Post-hoc* analyses of the significant effects were computed by means of the Newman-Keuls test, $p = 0.05$ or less.

The significant main age effect ($F_T(1,93) = 23.42$, $p < 0.0001$; $F_L(1,93) = 28.42$, $p < 0.0001$) revealed that young subjects performed faster than the elderly (68.98 *vs* 84.81 sec). The significant main effect of education ($F_T(2,93) = 3.674$, $p < 0.03$; $F_L(2,93) = 5.52$, $p < 0.005$) revealed that high level subjects performed faster than low level subjects (70.13 *vs* 82.35 sec; middle level = 81.61). The impressively significant main condition effect ($F_T(2,186) = 780$, $p < 0.0001$; $F_L(2,186) = 474$, $p < 0.0001$) showed that condition R (44.71) was performed faster than condition N (64.95) and condition N faster than condition I (119.82). However, these three main effects were qualified by the significant condition x age interaction ($F_T(2,186) = 6.497$, $p < 0.002$; $F_L(2,186) = 19.06$, $p < 0.0001$) which, in turn, was qualified by the significant age x education x condition interaction ($F_T(4,186) = 2.558$, $p < 0.04$; $F_L(4,186) = 3.17$, $p < 0.02$). The ANOVA on untransformed response times also revealed significant age x education ($F_T(2,93) = 3.85$, $p < 0.025$) and condition x education ($F_T(4,186) = 3.17$, $p < 0.01$) interactions.

The age x education x condition interaction is depicted in Figure 1 and the *post-hoc* analysis showed the following: an effect of education appeared only in conditions I and N for elderly subjects (favoring high over low and middle levels), and in condition R for young subjects (favoring middle and high levels over low level); an age effect, favoring the young subjects, appeared only for the low and middle levels in conditions I and N, as well as for middle level in condition R; and finally, the main effect of condition applied exactly to the six subgroups of subjects.

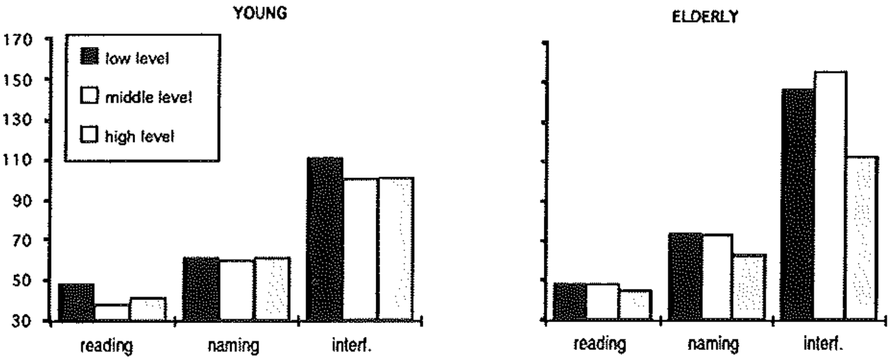

Figure 1: Mean response time (in sec) to read words (reading), to name colors (naming) and to name the color of color-words (interference) as a function of age and education level.

Measures of Stroop interference

As a matter of fact, the direct measure of interference is the subtraction of condition N from condition I (Stroop effect). In addition, it seems safe to weight this difference by the general individual efficiency of the subject. So, in order to investigate more specifically the effects of age and education on the Stroop effect, *an individual index of interference* was computed for each subject and studied by means of an age x education ANOVA. The index was computed as (I – N)/(I + N), and the limits of the resulting values were between – 1 (maximal «anti-Stroop effect») and + 1 (maximal Stroop effect). Only age reached a significant level: the interference was higher in elderly than in young subjects (0.100 *vs* 0.117: $F_L(1,93) = 5.604$, $p < 0.02$; $F_T(1,93) = 7.32$, $p < 0.009$).

However, it could be that this result was biased by a natural individual tendency to respond more readily to words or to colors. Therefore, *a response dominance index* was calculated (Panek *et al.*, 1984) by computing the individual difference between condition N and condition R and entering this score as a covariate in an additional analysis of covariance (ANCOVA) on the interference index, with age and education as between-subject factors. The previous result was verified, that is to say, only the main age effect was significant ($F_L(1,92) = 9.56$, $p < 0.003$). However, this effect disappeared when the untransformed data were considered ($F_T(1,87) < 1$).

Thus, on transformed response times, the effect of age on the index of interference was not an artifact of an age effect on response dominance. And indeed, an age x education ANOVA computed on the response dominance index (logarithmic transformation) revealed only a significant age x education interaction ($F_T(2,93) = 3{,}552$, $p < 0.033$), which was only due to a lower dominance index in young low level subjects than in the remaining five subgroups. Conversely, on untransformed response times, the effect of age on the index of interference was probably an artifact of an age effect on response dominance (however, the complementary age x education ANOVA of the dominance index did not reveal any significant effect).

Thirty years ago, Jensen (1965) detected in the literature no less than eleven different types of scoring the Stroop test, and compared them. In particular, three factors emerged from a factor analysis: the «color naming» factor scored as N (N-R), the «interference» factor measured as I–N, and the «speed» factor expressed by R (our nomenclature). Note that our index of interference is the «interference factor» weighted by the general performance of the subject. Houx *et al.* (1993) proposed another index of interference, computed as $\{I - [(R+N)/2]\}/[(R+N)/2]$. These four formulas were applied to the present data, and submitted to separate analyses of variance with age and education level as the two between-subject factors. Only the main age effect was significant for the «speed» factor ($F_L(1,93) = 7.520$, $p < 0.0075$; $F_T(1,93) = 5.881$, $p < 0.018$), for the «interference» factor ($F_L(1,93) = 7.044$, $p < 0.0095$; $F_T(1,93) = 18.73$, $p < 0.0001$) and for the interference as measured by the formula of Houx *et al.* ($F_L(1,93) = 7.225$, $p < 0.009$; $F_T(1,93) = 11.18$, $p < 0.002$). For the «color naming» factor, the sole significant effect was the age x education level interaction on transformed data ($F_L(2,93) = 3.237$, $p < 0.044$; $F_T(2,93) = 1.845$, NS), which was due to low level subjects where young subjects obtained a significantly lower value than the elderly. The sole additional significant effect was the education main effect on the «interference factor» for untransformed data ($F_T(2,93) = 4.13$, $p < 0.02$).

Discussion

The Stroop effect is evidenced by the shorter time for naming colors than for naming the color of color-words (MacLeod, 1991, for a review; Stroop, 1935). The present study verifies the Stroop effect, since the basic analyses – on both untransformed and logarithmically transformed times – bring it out clearly in the six subgroups of subjects designed by crossing two age levels and three education levels. In addition, the phenomenon cannot be considered as an artifact resulting from a simple advantage of naming over reading since an advantage of word reading over color naming was systematically observed, and since the interference effect was maintained when response dominance was taken as a covariate (at least on transformed data).

However, it could be that the Stroop effect – like many other cognitive functions – is affected by several subject variables, especially age and education (see Introduction).

Age

According to the sixteenth «major empirical result» listed by MacLeod (1991), Stroop interference begins early in the school years, increases as reading skill develops, then declines through the adult years until about the age of 60, at which point it begins to increase again. The recent data of Houx *et al.* (1993) supports this conclusion, since the interference effect was higher in subjects over the age of 60 (especially if they had suffered «biological life events»). This was verified in their direct measure of interference. Our results, too, support this conclusion, since reading time in the interference condition was higher in elderly (age 60-79) than in young subjects (age 18-35) and, accordingly, the direct measure of interference was higher in elderly than in young subjects, even when the response dominance was taken as a covariate, and even when the formulas of Jensen (1965) and of Houx *et al.* (1993) were used to measure interference. This greater Stroop interference in elderly adults is consistent with the view that age increases the difficulty of inhibiting irrelevant information (see Zacks & Hasher, 1994).

Education

Houx *et al.* (1993) showed that the Stroop effect was affected by education, an individual variable not considered in the review of MacLeod (1991). In this first study (to our knowledge) revealing an effect of education, the effects of age and of age x presence/absence of «biological life events» were amplified in subjects with a low level of education, and the direct measure of interference showed a significant main effect of the level. In the present study, the age effects reported above applied only to low (naming and interference) and middle levels (the three conditions) and, in addition, an effect of the education level appeared in the elderly subjects where, in the interference and naming conditions, subjects of high level performed faster than subjects of the remaining two subgroups; in the young subjects, there was an advantage of middle and high levels over low level in the reading condition only.

Thus, up to now, only two studies have made a contribution to the sensitivity of the Stroop test with respect to education, with slightly different results. In the study of Houx *et al.* (1993), less educated subjects performed more slowly than more highly educated subjects, the condition effect was amplified in less educated subjects, and the increased susceptibility to interference with age was increased in less educated subjects (an age x education level x condition significant interaction); in addition, the interference index was higher in less educated than in more highly educated subjects. In our study, less educated subjects performed more slowly than more highly educated subjects, but this was due to the interference and color naming conditions in elderly subjects, and to the reading condition in young subjects (an age x education level x condition significant interaction, again); however, these effects of education level disappeared when the interference index was considered, even when this index was computed according to the formulas of Jensen (1965) and of Houx *et al.* (1993), and even when the response dominance was taken as a covariate.

This comment raises the question of the kind of measure that is taken for evaluating the Stroop effect.

Kinds of measure of the Stroop effect

The Stroop effect is a robust one (MacLeod, 1991, for a review) and should not depend on the chosen measure of the dependent variable. However, the kind of measure could become more critical when the effect of some subtle subject variables on the Stroop interference is considered. This also applies to the choice of index to express Stroop interference. On the one hand, the logarithmic transformation is appropriate when the total time to perform a task is the dependent variable. In the present study, all analyses were computed on both the raw and the transformed data. For basic analyses, the same conclusions were reached from both kinds of data (with just two minor exceptions). Nevertheless, some discrepancy emerged for derived indices. Thus, caution is required when interpreting the results, and one has to be aware of these «technical» complications. On the other hand, the Stroop interference can be diagnosed by several means. In the present study, four measures were considered and other measures were possible (for instance $(I - N)/N$).

Fortunately, similar conclusions were reached. However, once again, caution is required when interpreting the results, and one has to be aware of this diversity of «technical» choices. Considering the general slowing that accompanies aging, it seems that a ratio score would be better than a difference score.

SUMMARY

The Stroop effect is evidenced by a shorter time for naming colors than for naming the color of color-words. It is well known that this effect is amplified by age, but little is known about the effect of education. In the present study, nearly one hundred subjects were submitted to the classical Stroop's (1935) paradigm. They were divided into six subgroups by crossing two age levels and three levels *of education. The classical Stroop effect was observed in all subgroups, and was amplified in the elderly as compared to young adults. The effect was not clearly affected by education, which is at variance with the sole other study on this topic (Houx, Jolles & Vreeling, 1993). However, the present study illustrates that the choice of the dependent variable (raw response times* vs *logarithmic transformation of the response times) and of the index of interference could be critical when the influence of individual variables on the Stroop effect is sought.*

RÉSUMÉ

Ce que l'on appelle « effet Stroop » se manifeste par un temps de dénomination des couleurs plus court que le temps de dénomination de la couleur dans laquelle sont écrits des noms de couleurs. S'il est bien établi que ce phénomène s'accroît avec l'âge, on connaît en revanche mal les effets du niveau d'éducation. Dans cette étude, une centaine de sujets adultes ont été soumis à la procédure classique de Stroop (1935). Les sujets constituaient six sous-groupes, par combinaison de deux niveaux d'âge et trois niveaux de scolarité. L'effet Stroop fut observé dans tous les sous-groupes et était plus important chez les personnes âgées que chez les jeunes adultes. L'effet n'était pas clairement affecté par le degré d'éducation, ce qui ne confirme pas la seule étude connue sur cette question (Houx, Jolles & Vreeling, 1993). Cependant, la présente étude indique que le choix de la variable dépendante (temps brut vs *transformation logarithmique de ce temps) et de l'index d'interférence pourrait s'avérer critique lorsqu'on cherche l'influence de variables individuelles sur l'effet Stroop.*

ZUSAMMENFASSUNG

Der sogenannte « Stroop Effekt » äussert sich durch einen kürzeren Zeitraum, den Vps benötigen für die Benennung von Farben im Vergleich zur Benennung der Farben von Farbennamen. Es steht zwar fest, dass sich dieser Effekt mit dem Alter vergrössert, aber der Einfluss von Ausbildung der Vps ist wenig bekannt. In dieser Untersuchung wurde etwa hundert Erwachsenen die klassische Stroop Prozedur vorgeschlagen. Die Erwachsenen wurden in sechs Untergruppen aufgeteilt durch die Kombination von zwei Altersstufen und drei Ausbildungsstufen. Der klassische Stroop Effekt wurde in allen Untergruppen beobachtet; er kam bei älteren Erwachsenen mehr zum Vorschein als bei jüngeren. Der Effekt wurde durch die Ausbildungsstufe nicht deutlich beeinflusst, wodurch die Ergebnisse der einzigen bekannten Untersuchung auf diesem Gebiet (Houx, Jolles & Vreeling, 1993) nicht bestätigt wurden. Dennoch weist die vorliegende Untersuchung darauf hin, dass die Wahl der abhängigen Variabel (Bruttozeit vs *logarithmische Umsetzung dieser Zeit) und der Interferenzindex sich als kritisch herausstellen könnten, wenn man den Einfluss individueller Variabeln auf den Stroop Effekt festzustellen sucht.*

REFERENCES

ARBUCKLE, T.A., GOLD, D., & ANDRES, D. (1986). Cognitive functioning of older people in relation to social and personality variables. *Psychology and Aging, 1*, 55-62.

ARBUCKLE, T.A., GOLD, D.P., ANDRES, D., SCHWARTZMAN, A., & CHAIKELSON, J. (1992). The role of psychosocial context, age, and intelligence in memory performance. *Psychology and Aging, 7*, 25-36.

COHN, N.B., DUSTMAN, R.E., & BRADFORD, D.C. (1984). Age-related decrements in Stroop color test performance. *Journal of Clinical Psychology, 40*, 1244-1250.

COMALLI, P.E., WAPNER, S., & WERNER, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology, 100*, 47-53.

CRAIK, F.I.M., BYRD, M., & SWANSON, J.M. (1987). Patterns of memory loss in three elderly samples. *Psychology and Aging, 2*, 79-86.

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

HOUX, P.J., JOLLES, J., & VREELING, F.W. (1993). Aging effects assessed with the Stroop color-word test. *Experimental Aging Research, 19*, 209-224.

JENSEN, A.R. (1965). Scoring the Stroop test. *Acta Psychologica, 24*, 398-408.

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

PANEK, P.E., RUSH, M.C., & SLADE, L.A. (1984). Locus of the age-Stroop interference relationship. *Journal of Genetic Psychology, 145*, 209-216.

SALTHOUSE, T.A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

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

VAN DER LINDEN, M., WYNS, CH., BRUYER, R., ANSAY, C., & SERON, X. (1993). Effect of educational level on cued recall in young and elderly subjects. *Psychologica Belgica, 33*, 37-47.

ZACKS, R.T., & HASHER, L. (1994). Directed ignoring. Inhibitory regulation of working memory. In D. Dagenbach & T.H. Carr (Eds), *Inhibitory processes in attention, memory and language* (pp. 241-264). San Diego: Academic Press.

First authors' address:
UCL-NECO
Place du Cardinal Mercier 10
B-1348 Louvain-la-Neuve