Introducing Pattern Separation (PS) is the process by which similar inputs are transformed into separate non-overlapping representations. It is a critical function of episodic memory and has a notable survival value, since it allows people to compare everyday situations to those previously encountered, thus eliciting an appropriate behavioral response. There are two main lines of research on PS in humans. One line of research has been developed based on the observation that hippocampal neurogenesis is involved in stress, mood, and fear regulation, and that impaired PS performance has been observed in individuals with subclinical depressive symptoms. The relationship between PS and difficulty recollecting details and recalling appropriate memories has been hypothesized to be disadvantageous for PS, thus disadvantaging the recognition of similar experiences as separate, non-overlapping representations. It is usually assessed via the Mnemonic Similarity Task – Object Version (MST-O) which, however, assesses PS performance without taking behavioral context discrimination into account, since it is based on pictures of everyday simple objects on a white background. We here present a validation study for a new task, the Mnemonic Similarity Task – Context Version (MST-C), which is designed to measure PS while taking behavioral context discrimination into account by using real-life context photographs.

Methods: Fifty healthy subjects underwent the two MST tasks to assess convergent evidence. Instruments assessing memory and attention were also administered to study discriminant evidence. The test-retest reliability of MST-C was analyzed.

Results: Weak evidence supports convergent validity between the MST-C task and the MST-O as measures of PS ($r_s = 0.464; p < 0.01$); PS performance assessed via the MST-C did not correlate with memory or attention; a moderate test-retest reliability was found ($r_s = 0.595; p < 0.01$).

Conclusion: The MST-C seems useful for assessing PS performance conceptualized as the ability to discriminate complex and realistic spatial contexts. Future studies are welcome to evaluate the validity of the MST-C task as a measure of PS in clinical populations.

Keywords: Pattern separation; context discrimination; mnemonic similarity task; Mnemonic Similarity Task – Object Version; Behavioral Pattern Separation Task
Unfortunately, the literature on PS in psychiatric disorders is still poor and lacks promising findings that can contribute to definitions of the neural correlates and mechanisms underlying these clinical features. Thus, syndromal models could be developed with multiple and complex etiologies and could include overlapping genetic and environmental risks.

PS performance has usually been measured with the Mnemonic Similarity Task – Object Version (MST-O), previously called Behavioral Pattern Separation Task – Object Version (BPS-O). The MST-O has been used in studies evaluating age-related decline of mnemonic discrimination, verifying the relationship between PS and anxiety in healthy individuals, and assessing PS in subjects with depressive symptoms. This is a computerized task that tests PS performance through a recognition memory paradigm involving colored photographs of simple everyday objects on a white background. The task seems to detect the specificity of hippocampal regions involved in different types of mnemonic discrimination. Indeed, functional magnetic resonance imaging (fMRI) research has shown that regions in bilateral CA3/DG were consistent with PS measured via the BPS-O. Bakker et al. found that activity in the bilateral CA3/DG in response to the presentation of a similar BPS-O item was significantly different from the activity in response to an old item. Similarly, Yassa et al. found a significant negative correlation between activity in the CA3/DG regions of the hippocampus and task performance, thus confirming a specific role for CA3/DG in PS. Moreover, the BPS-O was found sensitive in detecting specific age-related decline in mnemonic discrimination and in discriminating between MCI and unimpaired aging subjects.

Unfortunately, no studies are available on the validity or reliability of the MST-O as a measure of PS. Since it involves images of simple objects on a white background, the BPS-O does not seem fully adequate to manipulate and test context similarity, given that memories usually also include information on what is encountered in the context. In real-life, objects are presented within a rich, detailed mosaic of other features which creates a spatial context. Spatial context is part of the episodic memory and brain areas, such as the parahippocampal place area (PPA), are involved in processing global scene information. The fact that the MST-O does not take spatial context into account would seem an important limitation, since perceptual complexity affects short-term visual memory capacity, which, in turn, relates to an increased perceptual similarity between complex items deriving from a large amount of overlapping information. In addition, a complex stimulus requires greater cognitive resources than a simple one to be represented and maintained in the memory; thus, assessing subjects using simple stimuli might underestimate the performance of those with greater cognitive resources. Finally, a recent study by Libby et al. found that hippocampal neural activity changed according to item-context manipulation. In brief, paradigms using colored photographs of everyday simple objects on a white background only roughly assess PS performance, since they do not consider behavioral context discrimination. In this framework, we propose a new task for assessing PS performance, the Mnemonic Similarity Task – Context Version (MST-C), which is based on real, complex, contextual photographs and is thus able to optimize the MST experimental paradigm in terms of the salience and generalization of the results. We here present evidence on convergent and discriminant validity regarding the association between MST-C task scores, as a measure of PS, and other variables. Convergent evidence is the degree to which test scores are correlated with tests of related constructs. Discriminant evidence is the degree to which test scores are uncorrelated with tests of unrelated constructs. Test-retest reliability, i.e., the degree to which an individual will obtain different results on a retest, was also analyzed.

Methods

Participants

Fifty healthy volunteers (50% females and 50% males; mean age: 30.66±11.73 years, age range: 21-59 years) were recruited from the general population of the Florence (Tuscany, Italy) metropolitan area. The exclusion criteria were: current or lifetime diagnosis of psychiatric disorder, assessed with the Mini International Neuropsychiatric Interview (MINI); current MCI, i.e., a score less than 24/30 on the Mini Mental State Examination (MMSE) or less than 26/30 at the Montreal Cognitive Assessment (MoCA); current or lifetime neurological disorder; current use of psychotropic medications; or currently undergoing psychotherapy. These additional exclusion criteria were assessed by a supplemental set of previously used interview-based screening questions that also investigate the first-degree relatives’ history of psychiatric disorders.

The majority of the sample was college students (44%), followed by white-collar workers (32%), blue-collar workers (12%), and unemployed individuals (12%). Most had a bachelor’s degree (56%), 34% had a high school diploma, and 10% had a middle school diploma. The majority (88%, n= 44) had no family history of psychiatric disorders; among the remaining subjects (12%, n=6), one had a family history of panic disorder and five had a family history of major depressive episodes.

Procedures

After having provided written informed consent, the subjects were assessed by trained psychologists with the MINI, the MMSE, and the MoCA to verify eligibility. If eligible, they completed the Stroop test, the Immediate Visual Memory (IVM) test, the MST-O, and the MST-C. To avoid familiarity and learning biases, the MST-O and the MST-C were administered in a randomized counterbalanced order. After 3 months, the participants repeated the MST-O and the MST-C.
Instruments

The MINI\textsuperscript{37} is a brief structured interview extensively used as a diagnostic instrument for psychiatric disorders according to DSM-IV-TR criteria.\textsuperscript{44} Its reliability and validity are well established\textsuperscript{35}; in the original study, kappa coefficients for concordance were good or very good for most diagnoses (from 0.53 and 0.82). The MINI’s specificity was good for all diagnoses (kappa coefficients from 0.72 to 0.97), and its test-retest reliability was acceptable (kappa coefficients between 0.76 and 0.93). The Italian version showed good inter-rater validity (kappa values above 0.73) and acceptable test-retest reliability (kappa values from moderate to good agreement).\textsuperscript{46}

The Folstein MMSE\textsuperscript{38} is a 30-item tool assessing cognitive functions (i.e., attention and orientation, memory, registration, recall, calculation, language, and ability to draw a complex polygon). Its total score ranges from 0 to 30, with scores of 24 or higher indicating no cognitive impairment and scores under 18 indicating severe cognitive impairment.\textsuperscript{46} In a non-clinical community setting, the MMSE had a pooled sensitivity of 85.1\% and a specificity of 85.5\%.\textsuperscript{47} The sensitivity and specificity of the Italian version were 85.7\% and 90\%, respectively, in a clinical sample.\textsuperscript{49}

The MoCA\textsuperscript{39} is a screening instrument for MCI. It evaluates 11 cognitive domains: visuo-constructional and executive skills, naming, memory, attention, abstraction, delayed recall, orientation. The total maximum score is 30; a score of 26 or higher is considered as no MCI.\textsuperscript{39} The MoCA was found to have very good test-retest reliability ($r = 0.92$) and internal consistency (Cronbach's alpha = 0.83). The Italian version\textsuperscript{50} showed good convergent evidence with MMSE scores ($p = 0.486$; $p < 0.01$).\textsuperscript{51} High sensitivity and specificity, accuracy of 0.96 (95\% confidence interval), and good inter-rater, as well as inter-rater reliability with respect to the MMSE.\textsuperscript{52}

The Stroop Color and Word Test – Short Version\textsuperscript{42} assesses executive functions (i.e., selective attention, effectiveness of concentration, cognitive flexibility, processing speed, and the ability to suppress automatic responses and interference inhibition). It includes three phases: in the first and in the second phases, the subject is asked to read a series of colored words and name their color, with attention measured according to the number of mistakes. The third phase is similar to the previous phases, except the words are all names of colors; interference control is measured in terms of reaction time (i.e., interference effect in time) and the number of mistakes (i.e., interference effect in mistakes), and is obtained by subtracting the third phase performance from the sum of the first and second phase performance.\textsuperscript{42} The cutoff of interference measures, adjusted for age and sex, is 36.91 seconds for interference effect on reaction time and 4.23 for interference effect on mistakes; above these thresholds, the subject is considered impaired.\textsuperscript{42} The Stroop test has good test-retest reliability ($r > 0.80$).\textsuperscript{53} The Short Version has been evaluated in a normative Italian sample ($n=248$; mean age: 52.1 ± 19.56 years, range 20-89 years).\textsuperscript{42}

The IVM is part of the Mental Deterioration Battery\textsuperscript{43} and evaluates immediate visual memory. The subject is asked to identify, among four alternatives, figures previously presented for a short time. The maximum score is 22; scores under 13.8, adjusted for age and education, are considered a sign of memory impairment. Single tests, as well as the entire battery, discriminate normal controls from demen-
ted patients with an accuracy of 95\%.\textsuperscript{43}

The MST-O\textsuperscript{5} has a first encoding phase in which the subject judges whether 128 items are indoor or outdoor. Then, there is a recognition memory phase of 192 items in which the subject judges whether they are old, similar, or new. The results are reported as rate of correct responses for each category of items.

The Mnemonic Similarity Task – Context Version (MST-C)

We here describe the procedure followed to create the MST-C. First, 110 pairs of similar photographs of real indoor or outdoor settings were collected, along with 128 unrelated items. Similar images shared the same context (e.g., a street with cars and trees), but some details of background were different (e.g., car color or model or the location of trees) (Figure 1).

The degree of perceptual similarity of each pair of similar images was evaluated according to the literature\textsuperscript{54} in a normative sample ($n=70$; males: 40\%; mean ± standard deviation [SD]: 24.43 ± 2.97 years); thus, subjects were asked to express their opinion on the level of similarity of each pair of images based on a 10-point Likert scale (from 1 = different; to 10 = identical). Pairs with a score in the highest (>) 75 percentile) and in the lowest (< 25 percentile) range of perceptual similarity were removed, leaving 64 pairs of similar photos (32 pairs of indoor and 32 pairs of outdoor items) and 128 unrelated items (64 indoor and 64 outdoor). This first 256-item version of the MST-C was programmed in E-Prime 2.0 software\textsuperscript{55} and pilot tested with a sample of 20 subjects (males: 40\%; mean ± SD age: 22.75 ± 2.19 years). Based on the rate of correct responses per pair and on the suggestion that items may be best studied when whole experimental samples yield 50\% correct responses,\textsuperscript{56} 13 similar pairs correctly labeled by less than 25\% of the subjects and four pairs correctly labeled by more than 75\% were considered too difficult or too easy, respectively, and were deleted and replaced with new photographs. A second pilot test was run in a new sample of 20 subjects (males: 33\%; mean ± DS age: 25.25 ± 3.17 years) and, due to the rate of correct responses per pair, all items were retained. At this point, convergent evidence between MST-C and MST-O scores was preliminarily tested. The average rates of similar items correctly identified on the MST-C and MST-O were 39.9 ± 13.05 and 48.25 ± 21.24, respectively, and the difference between the two rates (Wilcoxon test for dependent samples) was not statistically significant. The PS scores and rate of correct responses for both tasks were significantly correlated ($r = 0.487$; $p < 0.05$ and $r = 0.550$; $p < 0.05$, respectively).

The final version of the MST-C (Figure 2) includes 128 contextual images (64 indoor and 64 outdoor). In the first phase, each item is presented on a computer screen for 2 seconds followed by 0.5 seconds of interval and the subject is asked to judge whether the items are indoor or outdoor by pressing the V key or the N key, respectively.
In the second phase (i.e., unexpected recognition memory test), the subject is asked to classify a second series of images as either old, similar, or new compared to the first series of items. Thus, 192 items are presented; 64 are repetitions of items shown in the first phase (i.e., old), 64 are new items, and 64 are altered versions of previously shown items (i.e., similar); these altered versions were obtained by modifying the original setting or by digitally altering the images using Picasa, PhotoFiltre, IrfanView and/or MS Paint software. The results are reported as rate of correct responses for each category of items.

Statistical analysis

For the MST tasks, the PS score was calculated using two measures: a) the rate of similar items correctly identified minus the rate of similar items misidentified as new (SIS-SIN); b) the rate of similar items correctly identified minus the rate of similar items misidentified as old (SIS-OIS). The recognition memory score was also calculated as the rate of old items correctly identified minus the rate of new items misidentified as old. The number of correct responses for each category of items (i.e., old, similar, new) and the type of errors (i.e., identifications of new items as similar; identification of similar items as old) were also calculated.

The sample was stratified according to the following variables: sex; presence vs. absence of current or past psychiatric disorders among first-degree relatives (12% with vs. 88% without); education (median = 13 years of education; high: > 13 years of education; low: ≤ 13 years); and age (median = 26 years old; high: > 26 years old; low: ≤ 26 years old).

The influence of sociodemographic (i.e., sex, age, education, employment) and clinical variables (i.e., family history of psychiatric illness, MMSE, MoCA, Stroop test, IVM, MST-O, and MST-C) on MST performance was evaluated with the Mann Whitney U test for non-normally distributed variables. The Wilcoxon test for dependent samples was used to compare the rate of correct or incorrect responses on the MST-O and MST-C.

Spearman-Brown coefficients were calculated to examine: a) convergent evidence between the MST-C and MST-O; b) discriminant evidence between MST-C scores and immediate visual memory and executive functions; c) test-retest reliability of the MST-C and MST-O. The analyses were adjusted for family history for psychiatric disorders since it influenced PS in MST-O and memory in MST-C. Subjects with a family history of psychiatric disorders showed significantly lower scores in recognition memory (24.59 ± 28.77 vs. 47.29 ± 14.12; p = 0.02) and PS performance (16.18 ± 14.47 vs. 37.01 ± 20.45; p = 0.01) than the rest of the sample.
All analyses were conducted in SPSS version 20.0 for Windows. Significance levels were set at \( p < 0.05 \), and the strength of Spearman-Brown coefficients were evaluated according to cutoffs found in the literature\(^5\) (i.e., \( r \) from 0.00 to 0.30 negligible correlation; \( r \) from 0.30 to 0.50 low; \( r \) from 0.50 to 0.70 moderate; \( r \) from 0.70 to 0.90 high; \( r \) from 0.90 to 1.0 very high correlation).

**Results**

The mean scores of MMSE, MoCA, and IVM were 29.46\( \pm 1.01 \), 27.84\( \pm 1.17 \), and 19.36\( \pm 1.18 \), respectively. Stroop mistakes were 1.76\( \pm 0.84 \), Stroop interference effect in time was 21.95\( \pm 7.25 \) seconds, and Stroop interference effect in mistakes was 1.76\( \pm 0.85 \). Table 1 shows the...
means and SD of MST-C and MST-O scores for each sociodemographic variable. PS performance was not influenced by sex, age, or education level in either MST task. However, considering the effect size, even though the p-value was not statistically significant, the sex difference in MST-O score, when measured with the PS (SIS-OIS) formula, deserves consideration since it was approximately 15 points. Females performed worse than males when the PS(SIS-OIS) formula was applied to the MST-O, but not when applied to the MST-C.

Table 2 shows Spearman-Brown coefficients between PS scores in the two tasks. A statistically significant positive correlation was found between PS scores on the MST-C and the MST-O. Neither MST-C nor MST-O scores were significantly correlated with performance on the Stroop test or with IVM scores. Moreover, PS performance measured via the PS(SIS-OIS) formula in MST-C and in MST-O was not correlated with recognition memory ($r_s = -0.011$; $p = 0.930$ and $r_s = -0.118$; $p = 0.416$, respectively), while PS performance measured with the PS(SIS-SIN) formula in MST-C was correlated with recognition memory ($r_s = 0.418$; $p = 0.003$).

The rates of correct responses in MST-O and MST-C showed statistically significant correlations for new ($r_s = 0.456$; $p = 0.001$) and similar items ($r_s = 0.384$, $p = 0.006$), but not for old items ($r_s = 0.215$; $p = 0.134$).

MST-C test-retest reliability was $r_s = 0.595$ ($p = 0.000$) for the PS scores SIS-SIN; $r_s = 0.579$ ($p = 0.000$) for the PS scores SIS-OIS; and $r_s = 0.556$ ($p = 0.000$) for recognition memory score. MST-O test-retest reliability was $r_s = 0.603$ ($p = 0.000$) for the PS scores SIS-SIN; $r_s = 0.543$ ($p = 0.000$) for the PS scores SIS-OIS; and $r_s = 0.593$; $p = 0.000$ for recognition memory score.

**Discussion**

According to the cutoff provided in the literature, the MST-C showed low convergent evidence with the MST-O, high discriminant evidence with the IVM and Stroop test, and moderate test-retest reliability. Concerning the convergent evidence, PS performance was weakly correlated in the two MST tasks. This result is consistent with the hypothesis that PS performance conceptualized as real-life context discrimination and PS performance conceptualized as simple object discrimination do not overlap and that MST-O is inadequate for measuring PS performance in context.

Concerning discriminant evidence, there was no correlation between MST-C PS performance and attention, interference control, or immediate visual memory. This result is consistent with Toner et al. who found no correlation between the Stroop test and rates of correct identifications of similar MST-O items, as well as with Stark et al. who found no statistically significant correlations between PS performance and digit span performance or intelligence quotient. Thus, PS performance seems independent of executive function and memory.

MST-C PS score and recognition memory score were not correlated when the PS(SIS-OIS) formula was applied, and they were poorly correlated when the PS(SIS-SIN) formula was applied. These findings are consistent with
metric properties: low convergent evidence, high discriminant validity evidence of the MST-C as a measure of memory. Thus, the ability to distinguish among similar items seems unrelated to memory performance.

Concerning test-retest reliability, MST-C PS performance was highly stable after three months. Although the literature suggests that age influences PS performance, it did not in the present study. This inconsistency might be explained by the relatively young mean age of our sample.

The effect size observed for sex difference in the MST-O should be further explored. It has been recently suggested that, in rats, there are sex differences in neurogenesis regulation and PS, as well as in hippocampal function and neural plasticity. However, this result should also be evaluated in light of the fact that the effect size disappeared when PS was measured with the MST-C. Thus, again, spatial context might play a key role in PS performance assessment.

Some limitations should be pointed out. First, the sample may seem small and not fully representative of the general population, since subjects were recruited by convenience sampling. However, a number of studies on computerized cognitive tasks have been conducted with samples of less than 50 individuals, and convenience sampling is a standard procedure in this kind of research. Second, although we measured the degree of perceptual and mnemonic similarity of each pair of similar items, we did not provide a corresponding gradient in behavioral output, as recommended in recent publications. However, the results of the current study may be used in the future to provide a mnemonic similarity gradient for similar MST-C items. Third, the potentially high overall degree of semantic similarity among MST-C items might have biased the response. It could be claimed that to validate assessment of behavioral PS performance, novel abstract images would be preferable to familiar items in order to correct for semantic effects. However, activation in hippocampal regions has been observed in studies employing familiar items and in those using abstract, novel items. In addition, this potential bias was overcome by considering MST performance as the rate of correct responses to a similar item minus the rate of similar responses given to a new item.

Despite these limitations, the MST-C can surpass current instruments since it is a measure of PS performance that considers spatial context and has known psychometric properties: low convergent evidence, high discriminant evidence, moderate test-retest reliability. Thus, the MST-C can be considered useful for assessing PS performance in healthy subjects in real-life contexts and might be useful in clinical settings as a proxy for episodic memory function. From a research point of view, the MST-C may facilitate the study of neural correlates of PS performance in neuroimaging research, as well as help provide insight into the underlying etiopathogenetic mechanisms of psychiatric disorders. It could also be helpful in clinical trials, once the learning effects are controlled, due to its good test-retest reliability. Having a prevalence of visual content, it is easy to translate and adapt into other cultures, thus multicentric investigations and transcultural research could be undertaken.

Studies of healthy older adults, clinical populations (e.g., patients with cognitive impairment or affective, psychotic, or obsessive-compulsive-related disorders) or other variables are needed to confirm these preliminary findings and determine whether PS performance, as assessed by the MST-C, is influenced by semantic similarities or by an inter-item perceptual similarity effect, as well as to test the validity evidence of the MST-C as a measure of memory.

Disclosure

The authors report no conflicts of interest.

References

13 Shelly DJ, Kirwan CB. A possible negative influence of depression on the ability to overcome memory interference. Behav Brain Res. 2013;256:20-6.