Title:

Psychophysiological Detection of Feigned Memory Complaints

Aim:

The original aim was to identify psychophysiological parameters (both oculomotor and pupil dilation) of malingering during the performance of a memory test (i.e., The Test of Memory Malingering – TOMM; Tombaugh, 1996).

Changes to the initial aim:

1- Pupil dilation was collected, but dropped from the final analyses, due to concerns about possible variations in the luminosity of the space (i.e., during the period of the study the eye-tracking equipment had to be moved twice), which can interfere with pupil dilations.
2- Additionally, possible effects of age and education on test performance under normal effort and malingering conditions were explored by using the standard version of TOMM (i.e., without eye-tracking).

Method:

- Participants

The original plan was to recruit 150 healthy subjects from the community and 50 dementia patients from Hospital Santo António's outpatient clinic. The healthy subjects would be divided in three groups of 50 subjects each (i.e., naïve malingerers, coached malingerers, and normal effort).

Changes to the Participants:

1- Due to unforeseeable challenges in recruitment and logistics, only 92 individuals participated in the eye-tracking study (72 healthy individuals and 20 neurological patients with cognitive complaints). The dementia group was replaced by multiple sclerosis patients with cognitive complaints. After data curation, only 74 individuals were included in the final analyses: 31 normal effort (NE1), 21 naïve malingerers (NM), 8 coached malingerers (CM), and 14 patients with multiple sclerosis (MS). As planned, the NM were be asked to feign memory problems and were instructed to simulate memory deficit in order to receive a retirement or disability pension, without further coaching. The CM were asked to feign memory problems, were instructed to simulate memory deficit in order to receive a retirement or disability pension, and receivef coaching about symptoms and test performance strategies. The NE1 and MS subjects were asked to perform TOMM-C with their full-effort.

2- The standard version of TOMM was applied to 121 healthy adults recruited in the community and 46 patients with diagnosis of an autoimmune disease with central nervous system involvement (AID group; which included 42 patients with multiple sclerosis and 4 with neuropsychiatric systemic lupus erythematosus) from Hospital de Santo António's Outpatient Clinic. Fifty-five healthy participants were asked to feign memory impairment without coaching (FMI group) and the remaining 66 healthy participants were asked to provide their normal effort (NE2 group) while performing the TOMM.

- **Procedures**

As planned, all participants performed the TOMM - one of the most widely used performance validity tests in research and clinical practice. TOMM is a forced-choice visual memory recognition test and the number of correct responses is the standard measure to discriminate between true memory impairment from malingering.

For TOMM-C, three behavioural measures were recorded: Number of Correct Responses, Total Response Time, and Median Response Time on Correct Responses. Three oculomotor measures were considered: % of Total Number of Fixation on "new" items, % of Total Fixation Time on "new" items, and % of Fixation Time on "new" items for correct responses.

Changes to procedures:

1- Pupil dilation was collected, but dropped from the final analyses, due to concerns about possible variations in the luminosity of the space (i.e., during the period of the study the eye-tracking equipment had to be moved twice), which can interfere with pupil dilations.
2- Unlike the initial proposal, Dementia Rating Scale-2 before group assignment could not be performed due to logistical limitations.
3- The standard version of TOMM (without eye-tracking recording) was applied to a separate set of subjects and the Number of Correct Responses was the only recorded measure.

Statistical analyses:

As planned, the oculomotor data was processed with MatLab and statistical data analyses was conducted with Statistical Package for Social Sciences (SPSS). Descriptive statistics were used for group characterization and nonparamentric tests were applied to test the hypotheses.
Changes to the statistical plan:

- Diagnostic statistics and logistic regressions were also used.
- Two machine learning approaches (i.e., Support Vector Machine and Random Forest) were attempted (in collaboration with Paulo Castro Aguiar e Ana Filipa Gerós do i3S) to assess if eye-tracking metrics would be an informative complement to TOMM-C behavioral responses.

Results:

- TOMM-C: Group Differences

The NE1 group and the NM group had similar demographic characteristics, namely sex, age, and education. Though, the NE1 group was younger than the MS group and had fewer years of education than the CM group. NM individuals were younger and had more years of education than MS patients; and were younger than CM participants. No group differences were recorded regarding sex.

- TOMM-C: Group Performance

The NE1 group produced more correct responses, had shorter total response time and median response time on correct responses, lower % of total number of fixations, % of total fixation time on "new" stimuli, and % of fixation time on "new" stimuli for correct responses in comparison to the NM group on all TOMM-C trials.
The MS group were slower to respond (i.e., total response time and median response time for correct responses) on all trials, but had similar number of correct responses. On both Trial 1 and Retention Trial, the % of total number of fixations, the % of total fixation time on "new", and the % of fixation time on "new" for correct responses was higher for the MS group than the NE1 group. The MS group only differed from the NM group on the number of correct responses. Neither response speed nor oculomotor measures differed between MS and NM participants.
NM and CM groups did not differ on any of TOMM-C behavioral and oculomotor measures. The comparison between NE1 group and CM group revealed significant differences on all measures, but only partially (i.e., not on all trials).

The best cut-off scores were identified to differentiate NE1 and NM participants, while setting the specificity at >90%. The total number of correct responses had a sensitivity and specificity of 100% on Trial 1 and on Retention trial; the sensitivity dropped to 93.5% on Trial 1. The total response time produced relatively low sensitivity rates (ranging between 33.3% and 52.4%). The % of Total Fixation Time on “New” on Trial 1 was the oculomotor measure with the best sensitivity and specificity balance (respectively 81% and 93.5%).
The defined cut-off scores were then used to measure the frequency of abnormal scores in the MS group and CM group. Multiple logistic regression analyses revealed that, while taking into consideration demographic characteristics, MS patients had higher odds of abnormal score than NE1 participants on Total Response Time and Median Response Time for Correct Responses at Trial 1 and Trial 2. MS patients also had higher odds of abnormal score at Trial 1 on the following oculomotor measures: % of Total Number of Fixation on “New”, % of Total Fixation Time on “New”, and % of Fixation Time on “New” for Correct Responses. No statistically significant difference on oculomotor measures was found between NE1 and MS participants on Trial 2 and Retention trial, when demographic characteristics were considered.

- TOMM-C: Machine Learning

We took advantage of machine learning methods and frame the question in an automatic classification context. Eye-tracking variables did not improve the quality of the classification (i.e., sensitivity and specificity) provided by the behavioral responses.

- Standard version of TOMM: Group Differences

Healthy participants assigned to the NE2 and the FMI conditions and the AID group had similar sex, age, and education.

- Standard Version of TOMM: Detection of Malingering

FMI and NE participants’ median scores on the TOMM Trial 1 were respectively 35 and 47. All FMI participants scored <42 on the TOMM Trial 1, whereas only 1/66 of the NE subjects scored at that level (sensitivity of 100% and specificity of 98.5%). FMI and NE participants’ median scores on the TOMM Trial 2 were respectively 37 and 49. All FMI participants scored <45 and none of the NE group scored at that level on the TOMM Trial 2. FMI and NE participants’ median scores on the TOMM Retention Trial were respectively 38 and 50. All FMI participants and none of the NE group scored <48 at that trial. The sensitivity and specificity of TOMM Trial 2 < 45 and TOMM Retention < 48 were 100%.

- Standard Version of TOMM: Demographic Predictors of Poor Malingering

In the FMI group, scores ≤25 percentile of the distribution of the sample were considered poorer malingering. FMI participants who scored ≤31 on the TOMM Trial 1 were younger and had more years of education than those that scored >31. TOMM Trial 2 scores ≤33 were related to younger age and more years of education. Scores ≤34 on the TOMM Retention Trial were associated with younger age and had higher education. When age and education were considered as covariates, only education remained statistically associated with poorer malingering.

- Standard Version of TOMM: Normal Variability

NE participants who scored <50 on TOMM Trial 1 (n=57/66), Trial 2 (n=34/66), and Retention Trial (n=27/66) had lower education. On TOMM Trial 2, a score <50 was also associated with older age. AID participants who scored <50 on TOMM Trial 2 (n=18/46) had fewer years of education than those who scored 50 on Trial 2. Scores <50 on TOMM Trial 1 (n=38) or Retention Trial (n=19) were not significantly related with age or education.

Discussion:

Study results revealed that both behavioral response (i.e., response accuracy and response time) and eye-fixation data can distinguish malingerers from non-malingerers in a computerized version of TOMM.

The accuracy measure produced the most robust diagnostic statistics. The cut-off scores identified in the present study for the number of Correct Responses and the classification accuracy were higher than in most studies in the literature (Martin et al., 2020). Healthy individuals feigning memory impairment significantly produced fewer correct responses on TOMM-C than the NE group. NM and CM performance approached chance level, especially on Trial 1 and Retention Trial. The accuracy results confirmed the ceiling effect (almost all subjects produced the maximum score) in healthy individuals with credible performance (Rees et al., 1998). The number of Correct Responses was similar between MS patients and NE healthy individuals on all trials. Furthermore, only accuracy clearly differentiated MS patients from NM participants. These results provide support to its use in clinical practice, namely in patients with MS, memory complaints and mild memory difficulties.

Response time differentiated NE participants from both malingering groups, confirming previous reports (Bolan et al., 2002; Kanser et al., 2019) that healthy simulators are slower to respond on TOMM. However, MS patients were also slower to respond than healthy individuals under NE condition and had similar latency to the NM group. These results raise the need for caution when applying response time as a performance validity measure in clinical populations, namely in MS which is known to produce processing speed deficits in most patients (Ruano et al., 2017).

The oculomotor recording of the NE group showed a familiarity preference (i.e., shorter fixation time on "new" stimuli), especially on Trial 1, whereas both malingering groups showed a novelty preference (i.e., longer fixations on "new" stimuli than on previously presented stimuli) on the three TOMM-C trials. The eye-fixation measure with best diagnostic statistics in differentiating NE from NM participants was % of Total Fixation Time on "New" (sensitivity of 81.0% and specificity of 93.5%). These findings are consistent with a recent study (Tomer et al., 2020), which revealed that simulators spent more time gazing at irrelevant (i.e., foils) than relevant (i.e., target) stimuli in another PVT - the Word Memory Test. In non-clinical samples, a novelty preference appears to be a marker of non-credible performance on PVTs that require forced-choice recognition. It has been suggested that visual disengagement (i.e., gaze aversion) may be used by simulators to attenuate visual input and thereby decrease the cognitive load that they may be experiencing while performing the test (Tomer et al., 2020). Forced-choice memory recognition PVTs (e.g., TOMM and Word Memory Test) share some resemblance with the visual-paired comparison (VPC) tasks, which were designed to measure infant recognition memory. Both typically involve a familiarization phase followed by a test phase. During the familiarization phase, the individual is presented with a set of visual stimuli. During the test phase, the familiarization stimulus is paired with a novel stimulus. On VPC tasks, the spontaneous eye-movements are recorded and the amount of time spent looking at each stimulus during the test phase is usually the primary dependent variable. A decreased attention to familiar patterns relative to novel ones (i.e., spending more time looking at novel images) has been observed in VPC applied to preverbal human infants (Fantz, 1964), human adults (Manns et al., 2000), and primates (Pascalis & Bachevalier, 1999). VPC may also elicit a preference for familiarity depending on the length of the retention interval (Bahrick & Pickens, 1995; Richmond et al., 2007). Unlike standard VPC tasks, forced-choice memory recognition tests require an explicit recognition instruction and the visual behaviour of healthy adults during the test phase has been shown to favour familiar stimuli (Richmond et al., 2007; Brooks et al., 2023). Both the preference for novelty and the preference for familiarity are usually interpreted as evidence of recognition memory, whereas null preferences can be interpreted as evidence of forgetting (Richmond et al., 2007).

MS patients exhibited a less evident familiarity preference on the eye-fixation data than the NE participants at Trial 1. It's unclear why 50% of MS patients had a % of Total Fixation Time on "New" above the cut-off. Both the preference for novelty and the preference for familiarity are usually interpreted as evidence of recognition memory, whereas null preferences can be interpreted as evidence of forgetting (Richmond et al., 2007). It is reasonable to speculate that MS patients were more alert to the possibility that novel stimuli might be relevant, because of their prior experience with neuropsychological assessments (for clinical purposes) that require recall and recognition of previously presented stimuli without prior warning.

No effect of coaching how to avoid detection of malingering was observed on any of the behavioural and oculomotor measures of TOMM-C. In other words, the performance of NM and CM participants did not differ, which may reflect lack of statistical power or resistance of the test to coaching (Jelicic et al., 2011). Larger samples are necessary to confirm these negative findings.

The simultaneous recording of both behavioural and oculomotor measures in one of the most widely used PVTs and the exploration of both healthy individuals under different conditions and a clinical sample are strengths of the study. Though, the non-inclusion of a sample with suspected clinical malingering and the small size of the studied groups limits the generalizability of the research findings.

The between group differences regarding demographic characteristics limits the interpretation of the study results. In a complementary study, using the standard version of TOMM, revealed that age and education may interfere with malingerers' performance. Poor malingering on TOMM trials (scores $\leq 25^{th}$ percentile) was related to younger age and higher education. Malingerers with ≤ 12 years of education scored closer to the cut-off for non-credible performance on TOMM trials. These study results raise the possibility of decreasing sensitivity in milder forms of malingering depending on the level of education of the individuals.

Recent studies with pupillometry have reported that pupil dilation can detect feigned cognitive impairment on TOMM (Heaver & Hutton, 2011; Patrick et al., 2021). However, the present study focused only on eye-fixations. Future studies should explore the possibility of combining pupil reactivity with eye-fixation pattern in the detection of deception.

Conclusions:

In sum, healthy individuals feigning memory impairment showed a distinct behavioral (i.e., fewer correct responses and longer response times) and oculomotor (i.e., longer fixation time on “new” stimuli) response pattern on a computerized version of TOMM, reflecting an increased effort to inhibit a natural response. The results from a clinical sample with a bona fide neurologic condition and memory complaints point to the limited application of response time and eye-fixation measures in real-life clinical situations.

Recommendation:

Future studies ought to explore the preference for familiarity / novelty in bona fide MS patients and in other clinical populations and to investigate their associations with standard measures of memory (both visual and verbal).

References:

Bahrick, L. E., & Pickens, J. N. (1995). Infant Memory for Object Motion across a Period of Three Months: Implications for a Four-Phase Attention Function. *Journal of Experimental Child Psychology*, *59*(3), 343–371. https://doi.org/10.1006/jecp.1995.1017

Bolan, B., Foster, J. K., Schmand, B., & Bolan, S. (2002). A Comparison of Three Tests to Detect Feigned Amnesia: The Effects of Feedback and the Measurement of Response Latency. *Journal of Clinical and Experimental Neuropsychology*, *24*(2), 154–167. https://doi.org/10.1076/jcen.24.2.154.1000

Brooks, G., Whitehead, H., & Köhler, S. (2023). When familiarity not novelty motivates information-seeking behaviour. *Scientific Reports*, *13*(1), 5201. https://doi.org/10.1038/s41598-023-31953-6

Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science*, *146*(3644), 668–670. https://doi.org/10.1126/science.146.3644.668

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory*, *19*(4), 398–405. https://doi.org/10.1080/09658211.2011.575788

Jelicic, M., Ceunen, E., Peters, M. J. V., & Merckelbach, H. (2011). Detecting coached feigning using the test of Memory Malingering (TOMM) and the structured inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, *67*(9), 850–855. https://doi.org/10.1002/jclp.20805

Kanser, R. J., Rapport, L. J., Bashem, J. R., & Hanks, R. A. (2019). Detecting malingering in traumatic brain injury: Combining response time with performance validity test accuracy. *The Clinical Neuropsychologist*, *33*(1), 90–107. https://doi.org/10.1080/13854046.2018.1440006

Manns, J. R., Stark, C. E. L., & Squire, L. R. (2000). The visual paired-comparison task as a measure of declarative memory. *Proceedings of the National Academy of Sciences*, *97*(22), 12375–12379. https://doi.org/10.1073/pnas.220398097

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, *34*(1), 88–119. https://doi.org/10.1080/13854046.2019.1637027

Martins Da Silva, A., Cavaco, S., Moreira, I., Bettencourt, A., Santos, E., Pinto, C., Gonçalves, A., Coutinho, E., Samões, R., Dias, C. C., Teixeira-Pinto, A., Da Silva, B. M., & Montalban, X. (2015). Cognitive reserve in multiple sclerosis: Protective effects of education. *Multiple Sclerosis Journal*, *21*(10), 1312–1321. https://doi.org/10.1177/1352458515581874

Pascalis, O., & Bachevalier, J. (1999). Neonatal aspiration lesions of the hippocampal formation impair visual recognition memory when assessed by paired-comparison task but not by delayed nonmatching-to-sample task. *Hippocampus*, *9*(6), 609–616. https://doi.org/10.1002/(SICI)1098-1063(1999)9:6<609::AID-HIPO1>3.0.CO;2-A

Patrick, S. D., Rapport, L. J., Kanser, R. J., Hanks, R. A., & Bashem, J. R. (2021). Detecting simulated versus bona fide traumatic brain injury using pupillometry. *Neuropsychology*, *35*(5), 472–485. https://doi.org/10.1037/neu0000747

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, *10*(1), 10–20. https://doi.org/10.1037/1040-3590.10.1.10

Richmond, J., Colombo, M., & Hayne, H. (2007). Interpreting visual preferences in the visual paired-comparison task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 823–831. https://doi.org/10.1037/0278-7393.33.5.823

Ruano, L., Portaccio, E., Goretti, B., Niccolai, C., Severo, M., Patti, F., Cilia, S., Gallo, P., Grossi, P., Ghezzi, A., Roscio, M., Mattioli, F., Stampatori, C., Trojano, M., Viterbo, R. G., & Amato, M. P. (2017). Age and disability drive cognitive impairment in multiple sclerosis across disease subtypes. *Multiple Sclerosis Journal*, *23*(9), 1258–1267. https://doi.org/10.1177/1352458516674367

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, NY: Multi Health Systems.

Tomer, E., Lupu, T., Golan, L., Wagner, M., & Braw, Y. (2020). Eye tracking as a mean to detect feigned cognitive impairment in the Word Memory Test. *Applied Neuropsychology: Adult*, *27*(1), 49–61. https://doi.org/10.1080/23279095.2018.1480483