PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Impairment of brainstem implicit learning paradigms differentiates multiple system atrophy (MSA) from idiopathic Parkinson syndrome
AUTHORS	Sommer, Martin; Von Lewinski, Friederike; Schwan, Michaela; Paulus, Walter; Trenkwalder, Claudia

VERSION 1 - REVIEW

REVIEWER	Lucio Marinelli, MD, PhD Assistant professor of neurology Institute of Neurology, DINOGMI, University of Genova, Italy I declare I have no competing interests related to this article.
REVIEW RETURNED	05-Jun-2013

GENERAL COMMENTS	The authors evaluate a group of patients with multiple system atrophy (MSA) using eyeblink classical conditioning (EBCC) and serial reaction time task (SRTT) in order to assess implicit learning. The data are retrospectively compared with normal controls and patients with Parkinson's disease (PD) and progressive supranuclear palsy (PSP). It is concluded that, like in PSP, patients with MSA are impaired in EBCC learning paradigm, while data obtained with SRTT are less informative, mostly because patients with MSA found the task very difficult. The work is interesting and well written, anyway a few considerations concerning SRTT interpretation should be raised. The authors state clearly that SRTT is an implicit learning paradigm, since many works support this interpretation. It must be anyway underlined that the involvement of declarative or explicit resources has been demonstrated (Bo, 2011; Moisello, 2009). The very authors here specify that SRTT is an implicit measure of motor skill, but also report in the results the "retrieval of sequence", related to the explicit knowledge of the sequence. SRTT should be therefore considered an experimental paradigm were implicit and explicit learning components are entangled and difficult to tell apart. The authors should then specify that SRTT learning also involves declarative resources and discuss the data accordingly. Considering that MSA patients often are impaired in frontal executive functions and also in spatial working memory (Robbins, 1992), the SRTT result about similar retrieval of sequence between MSA and controls is very questionable and should be discussed more prudently (page 16, lines 3-5). Bo J, Jennett S, Seidler RD. Working memory capacity correlates with implicit serial reaction time task performance. Exp Brain Res. 2011 Sep;214(1):73-81. doi:10.1007/s00221-011-2807-8. Epub 2011 Aug 2. PubMed PMID: 21809082. Moisello C, Crupi D, Tunik E, Quartarone A, Bove M, Tononi G,

	Ghilardi MF. The serial reaction time task revisited: a study on motor sequence learning with an arm-reaching task. Exp Brain Res. 2009 Mar;194(1):143-55. doi: 10.1007/s00221-008-1681-5. Epub 2008 Dec 23. PubMed PMID: 19104787; PubMed Central PMCID: PMC2804101. Robbins TW, James M, Lange KW, Owen AM, Quinn NP, Marsden CD. Cognitive performance in multiple system atrophy. Brain. 1992 Feb;115 Pt 1:271-91. PubMed PMID: 1559159.

REVIEWER	Dr. Kai Lutz Senior Researcher Neuropsychology University Zürich and Scientific Director Clinic for Neurology and Rehabilitation cereneo AG
REVIEW RETURNED	01-Jul-2013

THE STUDY	The questions answered "No" are addressed in their order of appearance: ----------Q: Is the overall study design appropriate and adequate to answer the research question? Q: Are the participants adequately described, their conditions defined, and the inclusion and exclusion criteria described? The answers to both questions are linked, therefore I take the freedom to combine the answers: There were two research questions: 1) is procedural learning impaired in MSA? 2) may procedural learning be helpful for the differentiation of Parkinsonian syndromes? Ad 1) to answer this question, a MSA-patient sample has to be compared to a (comparable) sample of healthy controls, regarding performance in a suitable procedural learning task. Two issues emerge in the present manuscript: First, comparability of the control group is not clearly demonstrated: Cognitive and emotional status cannot be directly compared due to inconsistent test procedures, the matching procedure ("age matched healthy controls") is not explained, ex- and inclusion criteria are not explicitly listed. Second, it is questionable whether ECC is representative for procedural learning. Usually one would argue that a pointing task, tracking task, or SRTT should be used. Justification of ECC should be more explicitly demonstrated. Ad 2) Since the research question is expressed in a vague manner, it is appropriate to be satisfied with a vague answer, which can be given with the present investigation. Nevertheless, comparability of the subgroups should be elaborated on (cf. above): How was the matching performed, are there remaining differences in relevant characteristics like UPDRS, age, cognitive status? How could the great heterogeneity of severity of disease (UPDRS ranging from 5 to 69 in MSA patients) have affected results?

	At a later stage, two MSA subgroups are mentioned: MSA-P and MSA-C. How those subgroups identified? I am surprised that they were tested against each other by means of ANOVA. I can't remember reading anything about this ANOVA in the statistics section. All statistical analyses should be described there. Has there indeed been an ANOVA calculated, testing a sample of three against a sample of 7 patients? Concerning the study design, the following questions remain: - Analysis of the learning blocks, described on pages 6 and 7, is unclear. How were blocks aggregated? Does "ANOVAs with 'block' (blocks1-6, CS only block)" (page 9, line 8) mean, that factor "block" had 7 levels: Blocks 1 to 6 and CS only? Pleas make factors and factor levels explicit for all ANOVAs calculated. - What can presentation of an UCS have to do with random blinks (page 6, line 23) - how does presentation of an "UCS only" block control for such blinks?! maybe this is a language problem: What do the authors mean by "random blinks" anyway? Similarly: What is the "independent learning" effect, for which trial 11 is supposed to test? Are these blocks analysed at all? In table 1, a maximal score of 6 can be given for "Autonomic" assessment, but only 5 symptoms are given in the table legend. ---------- Q: Are the methods adequately described? - see above concerning matching procedure. - description of EMG recording should include the following information: sampling rate? unipolar or bipolar recordings? - How are "EMG-bursts" defined? Is it Peak-to-Peak amplitudes? or simply maximum signal intensity within a specified time window? ---------- Q: statistical Methods? - Application of ANOVAs assumes that data are normally distributed. This has to be tested, e.g., using a Shapiro-Wilk or Kolmogorow-Smirnov test. ---------- Q: Is the standard of written English acceptable for publication? - A few examples of bad language: page 4, line 15: "The serial reaction time task (SRTT) is another established task for which the implicit measures of motor skill (reaction time and errors) were close to normal in IPD patients...". I suppose the authors want to say that implicit motor skill [...], as measured by SRTT, was close to normal in IPD patients...? page 12, lines 13-14: "Both groups detected an increasing amount of the sequence during the course of the experiment..." Do they mean an increasing number of items of each sequence? Or of the last sequence? Or that subjects remembered more items of the sequence in post block reproduction of the last 10 items? Generally, there are many unclear formulations, which make it hard to understand what the authors would like to say. I would prefer to work on these, once the question is clarified whether or not the manuscript is suitable for publication.

RESULTS & CONCLUSIONS	---------- Q: Are they well presented? Although post-hoc t-tests have been calculated, any (explicit) reference to them is missing in the results section. The reader has to guess that asterisks, indicating significance, refer to post-hoc t-tests. Please check for statements in the results section referring to post-hoc tests and indicate them appropriately. I suspect that there will be several test results which are not described in the methods section. This might prove relevant for the mentioned Bonferroni-correction of post-hoc t-tests. ---------- Q: Are the interpretation and conclusions warranted by and sufficiently derived from/focused on the data? Some statements, especially statements which obviously refer to post hoc t-tests, can not be derived from the data as presented.

VERSION 1 – AUTHOR RESPONSE

Responses to Reviewer 1.

We would like to thank Dr. Marinelli for the comments and thoughtful suggestions.

The authors evaluate a group of patients with multiple system atrophy (MSA) using eyeblink classical conditioning (EBCC) and serial reaction time task (SRTT) in order to assess implicit learning. The data are retrospectively compared with normal controls and patients with Parkinson's disease (PD) and progressive supranuclear palsy (PSP). It is concluded that, like in PSP, patients with MSA are impaired in EBCC learning paradigm, while data obtained with SRTT are less informative, mostly because patients with MSA found the task very difficult. The work is interesting and well written, anyway a few considerations concerning SRTT interpretation should be raised.

Thank you for these encouraging comments.

The authors state clearly that SRTT is an implicit learning paradigm, since many works support this interpretation. It must be anyway underlined that the involvement of declarative or explicit resources has been demonstrated (Bo, 2011; Moisello, 2009). The very authors here specify that SRTT is an implicit measure of motor skill, but also report in the results the "retrieval of sequence", related to the explicit knowledge of the sequence. SRTT should be therefore considered an experimental paradigm were implicit and explicit learning components are entangled and difficult to tell apart. The authors should then specify that SRTT learning also involves declarative resources and discuss the data accordingly.

We recognize this concern and address the potential explicit aspects of the SRTT in the revised version (page 7 line 17).

Considering that MSA patients often are impaired in frontal executive functions and also in spatial working memory (Robbins, 1992), the SRTT result about similar retrieval of sequence between MSA and controls is very questionable and should be discussed more prudently (page 16, lines 3-5).

To address this point we now weaken even further the interpretation of the SRTT by mentioning the spatial working memory limitations (page 17, line 19).

Responses to Reviewer 2.

We would like to thank Dr. Lutz for the detailed comments and thoughtful suggestions.

Q: Is the overall study design appropriate and adequate to answer the research question?
Q: Are the participants adequately described, their conditions defined, and the inclusion and exclusion criteria described?

The answers to both questions are linked, therefore I take the freedom to combine the answers:

There were two research questions:
1) is procedural learning impaired in MSA?
2) may procedural learning be helpful for the differentiation of Parkinsonian syndromes?

Ad 1) to answer this question, a MSA-patient sample has to be compared to a (comparable) sample of healthy controls, regarding performance in a suitable procedural learning task. Two issues emerge in the present manuscript: First, comparability of the control group is not clearly demonstrated: Cognitive and emotional status cannot be directly compared due to inconsistent test procedures, the matching procedure ("age matched healthy controls") is not explained, ex- and inclusion criteria are not explicitly listed.

We recognize the concern on whether the groups are really matched, and we acknowledge the clear limitations of retrospective matching (page 8 line 5-7). In the revised version, we listed the matching criteria (page 5, line 6).

Second, it is questionable whether ECC is representative for procedural learning. Usually one would argue that a pointing task, tracking task, or SRTT should be used. Justification of ECC should be more explicitly demonstrated.

Thank you for raising this point. We now state more clearly the advantage of the EBCC, which is considered not to depend on (manual) motor skills that are likely to be impaired in parkinsonian disorders. (page 4, line 14).

Ad 2) Since the research question is expressed in a vague manner, it is appropriate to be satisfied with a vague answer, which can be given with the present investigation. Nevertheless, comparability

of the subgroups should be elaborated on (cf. above): How was the matching performed, are there remaining differences in relevant characteristics like UPDRS, age, cognitive status? How could the great heterogeneity of severity of disease (UPDRS ranging from 5 to 69 in MSA patients) have affected results?

In the revised version, we improved the description of the matching procedure (page 5, line 6). We now test (page 9, line 3) and report (page 11, line 15) that there are no remaining differences with regard to UPDRS score and age.

We now mention the wide range of motor impairments in the MSA group as a limitation (page 17 line 21).

At a later stage, two MSA subgroups are mentioned: MSA-P and MSA-C. How those subgroups identified? I am surprised that they were tested against each other by means of ANOVA. I can't remember reading anything about this ANOVA in the statistics section. All statistical analyses should be described there. Has there indeed been an ANOVA calculated, testing a sample of three against a sample of 7 patients?

We omitted this subgroup analysis due to the small sample size, as inferred by this reviewer (omissions on page 12 lines 16-17 and page 13 lines 1-2).

Concerning the study design, the following questions remain:
- Analysis of the learning blocks, described on pages 6 and 7, is unclear. How were blocks aggregated? Does "ANOVAs with 'block' (blocks1-6, CS only block)" (page 9, line 8) mean, that factor "block" had 7 levels: Blocks 1 to 6 and CS only? Pleas make factors and factor levels explicit for all ANOVAs calculated.

We now specify the description of blocks by making factors and factor levels explicit for all ANOVAs (page 9 and 10, on many lines).

- What can presentation of an UCS have to do with random blinks (page 6, line 23) - how does presentation of an "UCS only" block control for such blinks?! maybe this is a language problem: What do the authors mean by "random blinks" anyway?

There is a baseline level of eyeblinks irrespective of the EBCC task. We inserted a UCS only trial in each block in trying to control how many spontaneous blinks happen to fall within the time window where we expect a CR after CS presentation. This is what we call “random blinks” We report them numerically (page 12, line 4).

Similarly: What is the "independent learning" effect, for which trial 11 is supposed to test? Are these blocks analysed at all?

We rephrased this by stating that we controlled for the persistency of the learning effect using a CS only trial at the end of each block, which we analyzed independently. We now specify in the methods section how we dealt with the CS only trials (page 10 line 13). Their results are described on page 12, line 11, and on page 14, line 21).

In table 1, a maximal score of 6 can be given for "Autonomic" assessment, but only 5 symptoms are given in the table legend.

Thank you for mentioning this. Urinary incontinence and urinary retention were counted separately. (page 23 line 24).

Q: Are the methods adequately described?
- see above concerning matching procedure.
- description of EMG recording should include the following information: sampling rate? unipolar or bipolar recordings?
- How are "EMG-bursts" defined? Is it Peak-to-Peak amplitudes? or simply maximum signal intensity within a specified time window?

We now specify the EMG sampling rate of 10 kHz (page 7 line 5), better describe the unipolar montage setting (page 7 line 4), and elaborate the definition of EMG bursts (page 10, line 1).

Q: statistical Methods?
- Application of ANOVAs assumes that data are normally distributed. This has to be tested, e.g., using a Shapiro-Wilk or Kolmogorow-Smirnov test.

We now state clearly that we tested all datasets for sphericity using Mauchly's test, and that we applied Greenhouse–Geisser correction when necessary (page 11, line 1).

Q: Is the standard of written English acceptable for publication?
- A few examples of bad language:
page 4, line 15:
"The serial reaction time task (SRTT) is another established task for which the implicit measures of motor skill (reaction time and errors) were close to normal in IPD patients...". I suppose the authors want to say that implicit motor skill [...], as measured by SRTT, was close to normal in IPD patients...?

Thank you for pointing out these suboptimal wordings, we tried to rephrase this (page 4, line 16).

page 12, lines 13-14:
"Both groups detected an increasing amount of the sequence during the course of the experiment..."

Do they mean an increasing number of items of each sequence? Or of the last sequence? Or that subjects remembered more items of the sequence in post block reproduction of the last 10 items?

Your latter assumption is what we wanted to say, we corrected the wording on page 14 line 1.

Generally, there are many unclear formulations, which make it hard to understand what the authors would like to say. I would prefer to work on these, once the question is clarified whether or not the manuscript is suitable for publication.

We would appreciate further suggestions for improvement of linguistic clarity.

Q: Are they well presented?
Although post-hoc t-tests have been calculated, any (explicit) reference to them is missing in the results section. The reader has to guess that asterisks, indicating significance, refer to post-hoc t-tests.
Please check for statements in the results section referring to post-hoc tests and indicate them appropriately. I suspect that there will be several test results which are not described in the methods section. This might prove relevant for the mentioned Bonferroni-correction of post-hoc t-tests.

We recognize this and now mention post-hoc tests, where adequate, in the results section (page 14, line 18 and line 25) and figure legend (page 25, line 4) after describing them in the methods (page 10, line 23, and page 11, line 3).

Q: Are the interpretation and conclusions warranted by and sufficiently derived from/focused on the data?
Some statements, especially statements which obviously refer to post hoc t-tests, can not be derived from the data as presented. See above.

We improved this point by stating more clearly what post-hoc tests were done and if they were Bonferroni-corrected for multiple comparison (page 10, line 23, and page 11, line 3.

VERSION 2 – REVIEW

REVIEWER	Dr. Kai Lutz Neuropsychology, Institute of Psychology, University of Zürich,
REVIEW RETURNED	31-Jul-2013

GENERAL COMMENTS	All issues raised before were adressed appropriately. It may be worthwhile checking the manuscript for any further language errors.