

Please, refrain from making NLP a pseudoscience

© 2012 Joost van der Leij

NLP & pseudoscience

Neuro-Linguistic Programming (NLP) is a controversial method for improving people's lives. Some call NLP a pseudoscience, but they themselves seem to lack the scientific standing to make such a judgement.¹ The easiest way to counter any accusations of NLP being pseudoscience is to make it clear that NLP is not scientific. There are many valuable activities that are non-scientific like, for instance, philosophy. Although NLP practitioners claim that NLP help people to improve their live and many get paid to do so, everything is fine, as long as these NLP practitioners work on a no cure, no pay basis where the client gets to decide whether he finds talking to the NLP practitioner has been worth paying for. This doesn't mean that these NLP practitioners won't make unscientific statements. But as long as they refrain themselves from calling this a science it is okay. People make unscientific statements all the time and a great deal of unscientific statements in religion for example seem to improve some people's lives. Nor is anything lost by saying that NLP is not scientific. Something classifying as not being scientific only means that it hasn't properly been researched scientifically yet.

There are those of course who think that it has been scientifically proven that NLP doesn't work. But this is a misrepresentation of research that has shown that there is no scientific support for certain claims made by NLP practitioners.² The lack of scientific support for claim A doesn't imply that there is scientific support for not A as evidence against one hypothesis doesn't imply evidence for a different hypothesis. It simply means that claim A, for now, is non-scientific. Besides the misrepresentations there is also the suggestion that such research has been flawed:

¹ Witkowski, Tomasz (2011) - Thirty-Five Years of Research on Neuro-Linguistic Programming. NLP Research Data Base. State of the Art or Pseudoscientific Decoration? in Polish Psychological Bulletin Volume 41, Number 2 / 2010 p. 58-66 - Witkowski is not a scientific researcher working at a university, but a writer and a member of the Polish Sceptics Club. Almost all publications from Sceptics Clubs are non-scientific nor peer reviewed. Furthermore, the NLP Research Data Base is not unequivocally supported by the field of NLP worldwide.

² Sharpley, C. F. (1984). Predicate matching in NLP: A review of research on the preferred representational system. Journal of Counseling Psychology, 31(2), 238-248.

“The 6 categories of errors found in the literature include lack of understanding of the concepts of pattern recognition and inadequate control of context, unfamiliarity with NLP as an approach to therapy, lack of familiarity with the NLP meta-model of linguistic communication, failure to consider the role of stimulus-response associations, inadequate interviewer training and definitions of rapport, and logical mistakes”³

All in all this makes NLP an interesting case for a conceptual analysis of pseudoscience, non-scientific and unscientific. Especially because Hollander, a prominent Dutch NLP practitioner, is claiming to have scientifically proven that NLP works.⁴ This article will argue that Hollander’s endeavour is exactly the kind of behavior that substantiates claims that NLP is pseudoscience and that he ought to refrain from such activities. I will first investigate what the characteristics of pseudoscience are and how it relates to non-science and unscientific. Based upon these findings the article of Hollander will be analysed to see how it measures up against the criteria for pseudoscience. We will see that it fits the bill unfortunately. Finally, I will conclude that it is better to refrain from such pseudoscientific activities as they are neither helping science nor the field of NLP.

Pseudoscience, unscientific and non-science

5

The most commonly used definition of pseudoscience is:⁶

(1)	it is not scientific, and
(2)	its major proponents try to create the impression that it is scientific.

³ Einspruch, Eric L.;Forman, Bruce D. (1985) Observations concerning research literature on neuro-linguistic programming. Journal of Counseling Psychology, Vol 32(4), Oct 1985, 589-596.

⁴ Hollander, Jaap - The effectiveness of NLP: Interrupted time series analysis of single subject data for one session of NLP coaching. Unpublished, but the paper has been submitted in 2010 to the Journal of Counselling and Psychotherapy Research. Article can be found at: http://www.iepdoc.nl/artikelen/270_NLPartikelpublishversionfinal.pdf - The article is referenced on the frontpage of the website of Jaap Hollander’s training institute as “scientific proof for NLP” The paper is co-authored by Oliver Malinowski who at the time of writing was a master student and whom I consider to be the junior researcher. For that reason I only refer to Hollander for he is the senior research of this paper as I don’t want to burden a student with the problems of pseudoscience. Update: After writing this paper the article in question has been removed as has the reference that it has been submitted to a scientific journal. Yet it is still referenced as scientific proof for the effectiveness of NLP at <http://www.iepdoc.nl/wat-is-nlp/> In case it has been rejected ignoring the peer reviewed rejection by still claiming scientific validity supports the claim that this is nothing but pseudoscience as it ignores one of the basic tenets of modern scientific research. The article in question can now be found at: <http://www.iepdoc.nl/2010/01/01/effectiveness-of-nlp-interrupted-time-series-analysis-of-single-subject-data-for-one-session-of-nlp-coaching-2/>

⁵ This part is in large based upon the article in the Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/pseudo-science/>

⁶ Hansson, Sven Ove (1996). "Defining Pseudoscience", Philosophia Naturalis, 33: 169-176.

Pseudoscience is distinctive from non-science and unscientific. Non-science is a subject matter that has not (yet) been investigated by scientists. A certain remark is unscientific if science has proven something that contradicts it. Many NLP practitioners make unscientific remarks. Nevertheless this has no bearing on the scientific status of NLP. So far there are no scientific proofs that contradict NLP practices. If only for the fact that even though NLP is no science, sound NLP practitioners hold the principle that nothing in NLP ought to contradict science and if something in NLP is found to contradict science it has to be removed from NLP. Even critics of NLP claim: "NLP may be seen as a partial compendium of rather than as an original contribution to counseling practice and, thereby, has a value distinct from the lack of research data supporting the underlying principles that Bandler and Grinder posited to present NLP as a new and magical theory"⁷. So NLP is not unscientific but non-science. It only becomes a pseudoscience once major proponents of NLP create the impression that NLP is scientific while NLP is not scientific. It is precisely what Hollander does with his "scientific proof" of NLP.

There are a number of criteria to recognize pseudoscience. Something might lack falsifiability. "A sentence (or a theory) is empirical-scientific if and only if it is falsifiable"⁸. But these criteria do not cover all pseudosciences as some are falsified by science and are thus unscientific and still continue to present themselves incorrectly as a science. Furthermore, it is important to take paradigm changes into account because normal science may think that a revolutionary science is not falsifiable or has even been falsified by bad science. Nevertheless, pseudosciences still differ from science: "particular failures did not give rise to research puzzles, for no man, however skilled, could make use of them in a constructive attempt to revise the [pseudoscientific] tradition"⁹. Basically, this means that failures within a field ought to give rise to interesting puzzles. Puzzles which will inspire other researchers to revise their theories or practices. One valuable addition is that science doesn't concern itself with a single research, but only with much larger research programs¹⁰.

Another approach to show the criteria that make something a science is look at the epistemic norms that make a science. These norms are:¹¹

- universalism, the idea that scientific claims should be impersonal and not depending on the qualities of the researchers.
- communality, the idea that scientists are part of a scientific community and not isolated

⁷ Sharpley C.F. (1987). "Research Findings on Neuro-linguistic Programming: Non supportive Data or an Untestable Theory". *Journal of Counseling Psychology*. Journal 34, (1): 103-107.

⁸ Popper, Karl ([1989] 1994). "Falsifizierbarkeit, zwei Bedeutungen von", pp. 82-86 in Helmut Seiffert and Gerard Radnitzky, *Handlexikon zur Wissenschaftstheorie*, 2nd edition München: Ehrenwirth GmbH Verlag.

⁹ Kuhn, Thomas S (1974). "Logic of Discovery or Psychology of Research?", pp. 798-819 in P.A. Schilpp, *The Philosophy of Karl Popper*, The Library of Living Philosophers, vol xiv, book ii. La Salle: Open Court.

¹⁰ Lakatos, Imre (1981). "Science and pseudoscience", pp. 114-121 in S Brown et al. (eds.) *Conceptions of Inquiry: A Reader* London: Methuen.

¹¹ Merton, Robert K. ([1942] 1973). "Science and Technology in a Democratic Order", *Journal of Legal and Political Sociology*, 1: 115-126, 1942. Reprinted as "The Normative Structure of Science", pp. 267-278 in Robert K Merton, *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago: University of Chicago Press. Descriptions taken from Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/pseudo-science/>

individuals

- disinterestedness, the idea that a researcher has nothing to gain personally from the results of his research.
- organized scepticism, the idea that scientist ought to remain skeptical unless there is overwhelming evidence to the contrary.

Finally, there are a number of practical approaches to distinguish pseudoscience from science. Most of the time these take the form of checklist. One such a checklist is:¹²

1. *Belief in authority*: It is contended that some person or persons have a special ability to determine what is true or false. Others have to accept their judgments.
2. *Nonrepeatable experiments*: Reliance is put on experiments that cannot be repeated by others with the same outcome.
3. *Handpicked examples*: Handpicked examples are used although they are not representative of the general category that the investigation refers to.
4. *Unwillingness to test*: A theory is not tested although it is possible to test it.
5. *Disregard of refuting information*: Observations or experiments that conflict with a theory are neglected.
6. *Built-in subterfuge*: The testing of a theory is so arranged that the theory can only be confirmed, never disconfirmed, by the outcome.
7. *Explanations are abandoned without replacement*. Tenable explanations are given up without being replaced, so that the new theory leaves much more unexplained than the previous one.

NLP made pseudoscience

To see whether Hollander's research and publication make NLP a pseudoscience we need to establish that he is a major proponent of NLP who creates an impression that his NLP research and publication are scientific while they are in fact not. Hollander is in fact a prominent proponent of NLP. He has his own Wikipedia page which describes him as a well known national representative of NLP in the Netherlands and writer about NLP. He is founder of his own training institute, Institute for Eclectic Psychology which is one of the biggest NLP training institutes in the Netherlands.

If his research and publication turn out to be not scientific he then creates the impression that it is scientific while it is not. Not only is there the claim on the frontpage of the website of his training institute that references his research as the scientific proof of NLP, but his article is also structured according to guidelines for scientific publications. His article has an abstract. It makes references to scientific like methodologies such as Outcome Rating Scale (ORS), p-values, SPSS ARIMA and SPSS ANOVA. They all sound like terribly impressive scientific methods, thereby creating the impression of science and scientific research. The article has a list of references which is formatted according to scientific standards. In his biography he presents himself as the head of the research department of the Institute of Eclectic Psychology. If his research and article turn out not to be science at least he has done everything in his power to create the impression of them being scientific. To stress this even more he has submitted the article to a scientific peer reviewed journal, but it has been a long time since he has offered his article for publication without it being published. There is no update on whether his article has

¹² Hansson, Sven Ove (1983). *Vetenskap och ovetenskap*, Stockholm: Tiden.

been rejected or what causes the delay in publication. It does give the first clue that something might be wrong with the science of the article.

The lack of sound science

The major claim in the article is: "Results show that 16 individuals (64%) improved concerning their mild psychological and/or social problems through the NLP treatment." 16 out of 25 participants in the experiment had statistically significant improvements on their Outcome Rating Scale (ORS) with their p -value <0.05 . That all sounds like regular science. But even the article already remarks that there are problems with the experiment most notably the lack of a control group. The article then continues to list a number of problems where the experiment can be improved to get even better results. But that seems to glance over the problem of no control group. The only thing the article states about this problem is: "First of all, we did not use a control group. In fact, in this study participants are only treated with NLP. Therefore we are unable to show how NLP relates to other psychological treatment methods."¹³ It seems that Hollander is only worrying about showing the effectiveness of NLP vis-a-vis other psychological treatments methods. Maybe with the suggestion that even in this experiment NLP scores better than most other psychological treatments. But this line of reasoning completely ignores the more important question: does the lack of a control group matter in regard to the scientific status of the main claim that 64% of the participants reported improvements. This is a correct objection because the use of control groups is part of the golden standard of scientific testing.

But the lack of a control group doesn't have to be an objection. Fisher¹⁴ argues that is possible to draw conclusions from a single hypothesis. What statistics entails, he says, is that you create a null-hypothesis and try to disprove the null-hypothesis. Statistics can never prove a hypothesis but can at best only disprove a hypothesis. All other statistical methods work with at least two hypotheses as there are many good arguments that Fisher is wrong. But for the sake of the argument we can use Fisher for this discussion so as to see if the article in question is scientifically sound if we also agree, again for the sake of the argument, that there is no need for a control group.

But even if we follow Fisher, we immediately run into problems. Hollander does not create a null-hypothesis. In fact he gives us no hypothesis at all. This is the first major breach of scientific protocol. Nevertheless, we can amend the article and provide the missing null-hypothesis. The null-hypothesis ought to be "NLP does not improve mild psychological and/or social problems". If we can disprove this hypothesis we have a strong argument that NLP does improve mild psychological and/or social problems or that some weird case of chance has happened. The way Fisher proposes to see whether a hypothesis is disproved has been to subject the hypothesis to a test of significance. These are the famous <0.01 or <0.05 values. A test of significance basically means that the chance of a given result from an experiment is extremely unlikely. If the null-hypothesis is only correct in less than 1 in 20 experiments this is taken as scientific evidence that the null-hypothesis is wrong. If the null-hypothesis is only correct in less

¹³ Hollander, Jaap - The effectiveness of NLP: Interrupted time series analysis of single subject data for one session of NLP coaching. Unpublished.

¹⁴ Fisher, Ronald (1935) - The Design of Experiments, New York: Hafner Press

than 1 in 100 experiments this is taken as strong scientific evidence that the null-hypothesis is wrong. This is exactly what Hollander is doing with the reported values of improvements. Only if the improvements of an individual have a chance of less than 1 in 20 to be random the result is counted as a success. That sounds very scientific and it probably is. Unfortunately, it is the wrong statistic to look at as the claim Hollander wants to be scientific is that 64% of the participants have shown improvement. The correct question is: what is the chance that 16 out of 25 people would shown improvement?

The easiest way to answer this question is to use a control group because then you can look at how the values of the control group develop over time and compare the two. But we don't have a control group. This is no problem for Fisher. He suggests to think of a model that will reflect the situation. The easiest model to mirror the experiment is to flip coins. Either a participant of the experiment has a significant improvement or not. Due to the lack of a control group we do not know what the rate is by which people improve on their own. If we use a coin flip we assume that this rate is 50%. It might be argued that this is the wrong statistics to work with and I will do so later and show what percentage we can use based on other arguments. But as a first model coin flipping will do.

If we take coin flipping as the model, following Fisher we want to know what the chance is that with a fair coin we get 16 or more heads out of 25 flips. As it turns out the chance is 11.47% significantly more than the 5% needed for scientific evidence and considerably more than the 1% needed for strong scientific evidence. Even if we accept that the experiment is without a control group and go with the model of coin flipping for the experiment instead, it still shows no evidence for the effectiveness of NLP.

Hollander might argue that using the coin flipping is unfair. But without a control group he can't. The burden of proof lies with him and that is exactly the reason why it is so important to have a control group. Nevertheless, I am more than willing, for the sake of the argument, to introduce a proxy for a control group. For instance we can use the control group for experiments testing the effectiveness of antidepressants. The experiment done by Hollander tests for mild psychological and/or social problems. This in itself is a problem. The article is unclear about what is considered as a social problem as it gives only one example. It could very well be that in order to solve a social problem the participant only needs to be give the appropriate procedure to solve the social problem. For instance, a homeless person has the social problem of not having a home nor an income. To solve this problem it could be enough to instruct him to go to some official who will make sure he gets a home and an income. Procedures like these are part of NLP, but they differ greatly from the procedure to help people improve their psychological problems. It could very well be that from the 16 reported successes 8 participants were successful due to NLP procedures for psychological problems and 8 participants due to NLP procedures for social problems. Such a division would be disastrous for the experiments. Hollanders reported success rate would drop to 32%, but the test of significance would go up to 97%! A far cry from the 5% level. Such a result would almost be evidence of the opposite, but this only shows the limit of Fisher's reasoning. A 32% success rate might be quite good if only we could compare it to a control group. But even if we only take the one example of the article

as a separate group we get 15 successes out of 25 coin flips, which has a 21,22% chance of happening random. Again this is far removed from any indication of scientific evidence. Partitioning is a well known problem in statistics and in this case the use of disjunctions is questionable. It especially raises doubt about how the participants were selected and whether this disjunction has been introduced in order to accommodate the participants. I will return to these issues later on.

For now, lets see if we can use the control group of research into the effectiveness of antidepressants. NLP is often presented as an alternative for the use of antidepressants. A depression is definitely a psychological problem and it is at least mild and can be severe. Research into depression often uses the "Hamilton Rating Scale of Depression" (HRSD). That way the severity of the depression can be measured which is of course much better than the vague notion of mild psychological problems. So in principal one can find research about the effectiveness of antidepressants for mild depressions. I have not done so, but just taken the success rate of a meta-analysis of all kinds of depression: "Most patients, 56% to 60%, responded well to active treatment compared with 42% to 47% for placebo."¹⁵ As it turns out our model of coin flipping was pretty good as 47% in the control group got improvements in their depression without any help. But even we if take the 47% and recalculate the random chance on 16 success out of 25 tries according to Fisher, the chance is, 6,63% and still to high to pass a test of significance. So even if we take the results of the experiment in the most favorable conditions, it is still no scientific proof.

Other Fisherian considerations

Fisher mentions a number of other considerations that we need to take into account before we can make scientific inferences from data. For instance we need to make a forecast. We first need to define what we are looking for. Fisher would be against data mining. First we need to establish our null-hypothesis and then do the experiment. We are not allowed to create a hypothesis after the experiment has been done to fit the data from the experiment. In this paper there is no mentioning of any hypothesis let alone a forecast. So in this respect the experiment is deficient because we cannot establish to what degree the formulation of experiment has been influenced by the results. For instance it is possible that the choice for research into *mild* psychological and/or social problems has only been added after the fact. Limiting yourself to only *mild* psychological and/or social problems is strange as NLP makes no distinction between mild and severe problems. NLP practitioners claim success in both cases. Maybe the original experiment was meant to study psychological and/or social problems and ran into problems with severe psychological and/or social problems so people with severe problems were left out. Or maybe the social problems part was added after the experiment had led to an increase of the success rate. We don't know and in science the burden of proof is on the researcher to make sure that considerations like these are tackled in the publication. Without them no one can establish whether the results are scientific results.

¹⁵ Arroll B, Macgillivray S, Ogston S, Reid I, Sullivan F, Williams B, Crombie I. (2005) - Efficacy and tolerability of tricyclic antidepressants and SSRIs compared with placebo for treatment of depression in primary care: a meta-analysis.

Fisher finds it also important that we exclude other relevant causes from the experiment. There is no indication in this experiment that this is the case. Nowhere does it explain where the participants come from. The research has not been done by a university but by a private company, a NLP training institute, owned by the researcher. Many people with mild psychological and/or social problems come to such institutes to look for solutions. But this is a pre-selected group. Many of these people have undergone other kinds of psychological treatments and were not satisfied with the results. Many of these people have done a lot of additional work in self help or self development. If participants have been recruited through contacts from the NLP training institute there might be a strong bias in favor of NLP with these participants. It could be that participating in the experiment is perceived by the participants to get some free training, a kind of gift and that they want to give something back and hence report better results than they would otherwise. All of this is unknown due to the fact that nowhere in the article it is discussed where the participants come from, how they were recruited and what they got out of it by participating.

To show the influence of pre-selection compare this experiment with one of the NLP coaching program I have ran from 2006 till 2010. In total 100 people signed up for this program which costed 495,- euro. Most people who signed up had severe psychological problems. for instance a phobia to drive on the freeway or a three year depression due to the passing away of a partner. Part of the program was that if people felt that it didn't work they could get a refund of 60% or 297,- euro's. Officially the refund period was only a couple of week from the start of the program but in practice anyone requesting a refund in the following months or even years would get a refund no questions asked. Out of the 100 people only 8 people requested a refund. The chance of that happening at random if we take the 47% placebo base rate is too small for my binomial calculator as it lies very near to zero. In Fisherian significant testing it is a strong result with the null-hypothesis being rejected with a chance of <0.01 .

Nevertheless, this is no scientific proof of the effectiveness of NLP, because this was a highly preselected group of people. These people were selected by a procedure where they first got a free report and some free online video demonstrating NLP with the clear instructions to only sign up for this course if they noticed some positive changes when they followed the instructions. Furthermore, there were a couple of people with no psychological problems, but who just wanted to do the program to learn what I was doing. Finally, even though it was very easy to get a refund as you only had to ask for it, I didn't encourage people getting refunds for obvious reasons. So there are probably a few people who could have gotten a refund, but didn't request it. This shows how important it is for science to describe any experiment in clear detail.

Another important issue with Fisher is randomization. Nowhere in the article anything about randomization is mentioned. This could have led to the situation where the NLP practitioner and the participant knew each other personally or maybe they have done some other training together. In short, without randomization the list of other relevant causes for the reported success rate grows significantly.

To conclude, besides the experiment failing a test of significance, it also falls short on Fisher's

requirements for a forecast, excluding other causes and randomization. In short it is a bad experiment.

Beyond Fisher

There are considerably failings in Fisher's work with only one hypothesis. No-one is doing statistics that way anymore. All the other possibilities would work with two hypotheses, or in this case, with a control group and then compare those results.

The option most followed in the field of statistics is the Neyman-Pearson Theory (NP). In NP we have at least two hypotheses which we compare. NP theory focuses on two types of error people can make. Type 1 error is rejecting a hypothesis while it is true and a type 2 error is accepting a hypothesis while it is false. These two problems are also known as false positives and false negatives. Nowhere in the article are these error rates discussed.

But how important are these errors? Quite important according to Mayo: "We argue that the relevance of error probabilities is to ensure that only statistical hypotheses that have passed severe or probative tests are inferred from the data."¹⁶ Just showing that the data passes a 5% significance level or a p-value of < 0.05 is not enough. The experiment has to be set up in such a way that there are severe tests to investigate possible errors. Nothing of the sort has been done in this particular case. That Hollander is not very strict anyway is noted by the fact that he counts a participant with a p-value of 0,054 as a success. This is scientific misconduct. P-values of < 0.05 are already of lesser importance than p-values of < 0.01 and this way the test becomes less severe. Given the fact that with one success case less the chance of it being a random occurrence shoots up from around 10% to around 20%, having a severe test is of the utmost importance. We have already seen that by combining psychological and social problems there is at least one case in doubt and now we find another one.

And we still haven't look at possible errors. Of course with a test that asks for subjective scoring according to the Outcome Rating Scale, there is a lot of room for errors. It is well possible that someone has scored himself subjectively in a way that if questioned whether that was the right score, he would change his opinion and say: "No, you are right. I did in fact feel better or worse than I scored." Or people could forget to score them during the day and then just fill in the assessment based on memory, a lot less reliable than scoring each and every day. Working with something like the "Hamilton Rating Scale of Depression" works much better in that regard.

Of course we don't know what has happened in reality, but my point is that the burden of proof that everything went okay is on the researcher not on the reader. In fact if we look at participant #17 you can see that there is little margin for error. With a p-value of 0,041 it is close to the selected significance level of 0.05. With only one or two higher scores and all other scores pretty much around the same level only a small error might have pushed the p-value of participant #17 over the 0.05 level. This leaves us with already 3 possible errors. One error has happened for sure: a p-value of 0,054 has been treated as if it were a p-value of $< 0,05$. Whether

¹⁶ Mayo, Deborah and Aris, Spanos (2006) - Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction, British Journal for the Philosophy of Science 57(4), page 323-357

the second error (the disjunction of psychological and social problems) or the third error (scoring errors) have had an influence we don't know. But taking into account the known errors and the possible errors the number of success drops from 16 to 15 out of 25 for certain and maybe even to 14 out of 25. With the placebo rate of 47% in the depression case the chance of having 14 out of 25 people being helped as a random occurrence rises to 24.14% or almost 1 in 4 times people would do this experiment the same result would occur purely as a random event.

Also introducing Bayesianism or Likelihoodism in the hope of salving the experiment is no use. If we take our prior to be the placebo rate of 47%, the new data of 64% only gives us posteriors of 61%. Some Bayesians¹⁷ suggest that the strength of the evidence is the difference between the posterior and the prior, or in this case $61\% - 47\% = 14\%$. Hardly the material science is made of. There rests us no other conclusion than that there is at least bad science in the article by Hollander. This leaves the question if it is only bad science or whether it is actual pseudoscience.

Hollander and pseudoscience

Of course all this could be bad science. On the other hands it is quite easy for critics of NLP to take this article as an example demonstrating that NLP is a pseudoscience. The question is: are they right to do so? I fear they are for the following reasons. What I will do is go over all the different criteria that we found when discussing pseudoscience in general and see how they apply to this particular case.

The first important point is about falsifiability. Although Hollander has not mentioned the null-hypotheses nor is he talking about falsifiability, he is doing an experiment. Probably more to verify NLP than to falsify it, but this can be due to bad science rather than pseudoscience.

The second point is whether failures have let to research puzzles that could be used to revise NLP. While some practitioners in the field of NLP do regard failures and revise NLP based on those failures, Hollander is not one of them. In fact in a recent article Hollander has been defending an old failing version of NLP against exactly such a revision based upon previous failures¹⁸.

The third point is whether Hollander's research and article are part of a larger research program. Hollander might even claim this is the case as he presents himself as the head of the research department of his company. But in fact this is only a matter of appearances. There is no real research department to speak of at his company. Not in a way that can be compared to a research department at a university of major company. In fact it is more Hollander's private study that he counts as the research department. There is no building, no facility and no larger research program. If he would have presented himself as the owner of a training institute nothing would have been lost scientifically. By presenting himself as the head of the research department he is trying to create the impression that it is all science while in fact it is not. This is

¹⁷ Curd, Martin in Royall, Richard (2004) "The Likelihood Paradigm for Statistical Evidence" in Taper, Mark and Lele, Subhash "The Nature of Scientific Evidence: Statistical, Philosophical and Emperical Considerations" Chigago, University of Chigago Press - page 140-145

¹⁸ Hollander, Jaap (2011) "Modelleren: als de één het kan, kan de ander het leren!" - INzicht tijdschrift voor NLP jaargang 13 nummer 2

very typical of pseudoscience.

The fourth point is universalism. By using practitioners in training Hollander has taken this into account. A point for science.

The fifth point is communality. The research of Hollander is standing all on its own and is in no way part of a scientific community. Hollander may dispute this and claim ties to the local university, but these ties have nothing to do with a research program and are personal ties, not scientific ties. There is no peer reviewed published science article that mentions Hollander's research. Hollander has offered the article for publication but so far the magazine has not published it. Another point for the article to be pseudoscience.

The sixth point is disinterestedness. On this point Hollander fails miserably. Not only would any scientific proof of NLP let him to personal financial gains, but if it were he himself to proof this, it would make him even richer. In fact he is touting his "scientific proof" on his website and urges people to buy his trainings because he has "scientifically proven" NLP. Pseudoscience at its worst.

The seventh point is organized scepticism. Again Hollander fails to meet this point. There is no question in his mind whether NLP works or not. It works without any doubt. It is just a matter of proving scientifically that what he knows all along to be the case. This has nothing to do with any scepticism let alone it being organized. If that would have been the case, the experiment would be set up quite differently with attention to errors, better statistics and a control group.

The eighth point is belief in authority. Even though Hollander makes a lot of references to non-scientific resources for NLP and General Semantics, he doesn't claim any preferred status as an authority whereby he is right on all accounts. So this one goes to bad science.

The ninth point is nonrepeatable experiments. The experiment such as Hollander set up can easily be repeated. It is even simple to improve and run it again. So this point goes to science.

The tenth point is handpicked examples. We cannot answer this point as it is unclear how Hollander has selected the people he did the experiment with. If it was a random process this point goes to science. If they were pre-selected people without mentioning it, this point would go to pseudoscience.

The eleventh point is unwillingness to test. Hollander has obviously shown a willingness to test. The question remains though whether he is willing to accept a higher standard for his testing. So until that question has been answered, this point remains undecided.

The twelfth point is disregard for refuting information. On this point Hollander is guilty of pseudoscience. It is extremely easy to find articles scientifically criticizing NLP as they are for instance published on the Wikipedia entry for NLP. But there is no mentioning whatsoever by Hollander in his article. In fact he doesn't mention any criticism at all nor does he handle any critique. With a field so controversial as NLP this is a big omission and something that has not happened by accident. So it has nothing to do with bad science, but is a deliberate move that makes his article pseudoscience.

The thirteenth point is built-in subterfuge. With Hollander adding a participant with a p-value of 0,054 to count to his success rate this is a dubious point. It could be that it is just bad science. But the real question is: would Hollander have written and tried to get the article published if it turned out that it did not, according to him, proof the effectiveness of NLP? I cannot answer that question, but the whole atmosphere surrounding the article and the research leaves doubt that this would be the case.¹⁹

The fourteenth point is that explanations are abandoned without replacement. With no failures and no revisions this point is moot as Hollander has not been in the position where this would have been required.

To conclude: according to one point (universalism) Hollander does real science, on two accounts (falsifiability and authority) what he does is bad science, on five points it is undecided and according to the remaining six points what he does is pseudoscience. That is a lot of pseudoscience. Especially if you weigh each point. The points about revision, larger research program, communality, disinterestedness and organized scepticism are major points that outweigh many of the others. So there can be only one conclusion: retract the article and stop making NLP into a pseudoscience.

¹⁹ Update: since writing this paper the reference of the article in question being submitted for publication in a peer reviewed journal has been removed. Most likely due to it being rejected. If this in fact true this is another example of built-in subterfuge.