

# How You Say Or What You Say? Neural Activity in Message Credibility Evaluation

Lukasz Kwaśniewicz<sup>1</sup>, Grzegorz M. Wójcik<sup>1</sup>, Andrzej Kawiak<sup>1</sup>, Piotr Schneider<sup>1</sup>, Adam Wierzbicki<sup>2</sup>

<sup>1</sup> Maria Curie-Skłodowska University, Chair of Neuroinformatics and Biomedical Engineering, ul. Akademicka 9, 20-033 Lublin, Poland, [gwojczik@live.umcs.edu.pl](mailto:gwojczik@live.umcs.edu.pl)

<sup>2</sup> Polish-Japanese Academy of Information Technology, ul. Koszykowa 86, 02-008 Warsaw, Poland

**Abstract.** In the Internet era, understanding why humans find messages from unknown receivers credible (such as fake news) is an important research topic. Message credibility is an important theoretical aspect of credibility evaluation that relies only on message contents and design. For the first time in the field, we study message credibility by directly measuring brain activity of humans who make credibility evaluations in an experiment that controls message design. Brain activity as measured using EEG is used to investigate areas of the brain involved in message credibility evaluation. We also model and predict human message credibility evaluations using EEG brain activity measurements.

## 1 Introduction

“It is not what you say, but how you say it.” This proverb summarizes our common sense knowledge about persuasive communication. Philosophy, eristics, as well as modern marketing studies contain a wealth of knowledge on how to design a persuasive message. Unfortunately, such knowledge can be misused, as in the case of a phenomenon that has a huge detrimental effect on modern society, i.e. “fake news”. Proliferation of fake news today exploits the Web, which has been designed for a fast, cheap and easy publishing of information, but does not have adequate mechanisms that would support evaluation of information credibility. Fake news are, unfortunately, widely accepted by Web users who are not sufficiently critical or able to evaluate credibility of Web content [2]. Fake news are a subject of active research, but are still poorly understood [13].

The situation of a communication of (potentially) fake news can be abstracted into a model based on information theory. Consider Bob (the source of the message) who sends a message to Alice (the receiver): “If you have a headache, put some aromatic oil on your temples. Do not take aspirin, because it has bad side effects.” For simplicity, let us consider that the message only contains textual content. While Bob could communicate the message face-to-face (in this case, Bob’s looks and demeanor would probably strongly influence Alice’s reception of the message), but he could also use an e-mail, tweet, or Messenger message.

Upon receiving the messages, Alice must make the decision whether she believes the message or not. This decision is based on the credibility of the communicated message (which can be affected by several message properties, such as its persuasiveness) and on Alice’s knowledge concerning the subject. Researchers could simply ask Alice whether she believes the message. This was the approach used in social psychology [9, 20] that has studied factors that affect credibility evaluations, such as confirmation biases, group polarization, overconfidence and statistical illiteracy [9]. However, Alice may not want to answer such a question (for example, if the message concerns a sensitive subject), or be unable to give reasons of her credibility evaluation. Understanding these reasons could have important practical implications; for example, how to best design messages and information campaigns that correct and counteract fake news?

We hypothesize that credibility evaluation creates a mental state in Alice’s brain: Alice decides whether or not Bob (in her opinion) told her the truth. This mental state can be observed and measured using electroencephalography (EEG). Observing Alice’s brain activity could enable researchers to create an EEG-based measure of credibility and to better understand the factors (even unconscious ones) that influence Alice’s credibility evaluation.

However, most research that used EEG or fMRI in the context of credibility focused on lie detection [26, 15]. In other words, it studied Bob’s brain signals, not Alice’s. No attempts have been described in the research literature to investigate message credibility using EEG, which is the main subject of this article.

### 1.1 Research Problem and Contributions

The goal of this article is to address the following research questions:

- What brain areas are active when a receiver is evaluating message credibility?
- Does brain activity during credibility evaluation depend on message design?
- Can we model and predict human message credibility evaluations using EEG brain activity measurements?

One of the difficulties in addressing these question lies in the fact that experiment participants can have different levels of knowledge about the message. (In the above example: does Alice know the side effects of aspirin, or not?) An experiment for studying message credibility must ensure that participants have the same knowledge, and control the factors that may influence message credibility evaluation. In this article, we describe a pilot experiment that mimics the situation when the receiver has very little knowledge about the message subject, and can be influenced by irrelevant factors of message design. This situation reflects the reality of many Web users who encounter fake news on various subjects.

While the full answer to the questions listed above would require additional studies, this article makes significant contributions to the subject. We propose an experiment for studying message credibility evaluations using EEG. For the first time in literature, areas of the brain (Brodmann areas - BA) involved in the decision making based on message credibility are described in this article. Based

on this knowledge, the article describes an operational model of decision making based on message credibility that uses EEG measurements as an input. Not only does this model provide basic knowledge in the area of neuroinformatics, but it can be seen as a first step towards a practical EEG-based measurement method of message credibility.

In the next section, we introduce a definition of message credibility and discuss theoretical research that can guide the design of empirical experiments for studying credibility. We also discuss related work that studied brain activity related to credibility evaluation. In Section 3, we describe the design of our experiment. Section 4 discusses the experiment results. Section 5 concludes the article and introduces our plans for future work.

## 2 Related Work

### 2.1 Source, Message, Media Credibility

The concept of credibility, similarly to the concept of trust, is grounded in science and in common sense. Credibility has been subject to research by scientists, especially in the field of psychology and media science. One of the earliest theoretical works on credibility dates back to the 1950s. This influential work of the psychologist Carl Hovland [10] introduced the distinction between *source, message, and media credibility*. Out of these three, two are a good starting point for a top-down study of the complex concept of credibility: source credibility and message credibility. These two concepts are closely related to the natural-language definitions of the term “credibility”. In the English language dictionary (Oxford Advanced Learner’s Dictionary), credibility is defined as “the quality that somebody/something has that makes people believe or trust them”. When this definition is applied to a person (“somebody”), it closely approximates source credibility – an essential concept in real-life, face-to-face communication. However, notice that the dictionary definition of credibility can also be applied to “something” - the message itself. In many online environments, message credibility must be evaluated without knowledge about the source.

Information scientists have studied credibility evaluations with the goal of designing systems that could evaluate Web content credibility automatically or support human experts in making credibility evaluations [14, 27]. Credibility evaluations, especially of source credibility, are significant in online collaboration, for example on Wikipedia [25, 29]. However, human credibility evaluations are often subjective, biased or otherwise unreliable [18, 12], making it necessary to search for new methods of credibility evaluation, such as the EEG-based methods proposed in this article.

### 2.2 Message Credibility

A search for the term “message credibility” on Google Scholar returns over 1000 results (for an overview of recent publications, especially on the subject of Web content credibility, see [28]). Researchers from the media sciences have attempted to create scale for declarative measurements of message credibility

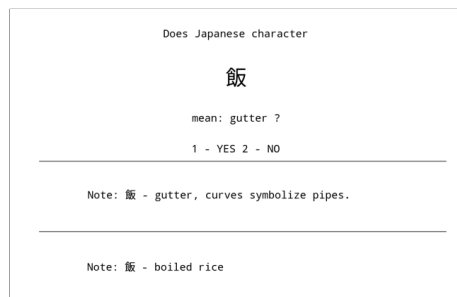
[3]. The importance of message credibility on social media has been recognized in many studies [28], for example in the area of healthcare [6].

As defined by Hovland, message credibility is the aspect of credibility that depends on the communicated message, not on its source or the communication medium. As such, message credibility depends on all information contained in the message itself. Consider a Web page that includes an article. The entire Web page is (in the information-theoretic sense) a message communicated to a receiver. Message credibility can depend on the article’s textual content, on images or videos embedded in the article, on Web page design and style, or even on advertisements embedded in the Web page.

This simple example shows that message credibility can be affected by many factors, or features of the message. Even if we limit ourselves to just the textual content of the message, message credibility is affected by both the semantic content of the message (its “meaning”) and by the pragmatic content of the message (its style, persuasiveness, sentiment, etc.) This is especially important since message credibility is usually evaluated rapidly. The work of Tseng and Fogg [24] introduced the two concepts of “surface credibility” and “earned credibility”, both of which can be applied to message credibility. Surface credibility is the result of a fast and superficial examination of the message. Earned credibility is the result of a slower and more deliberative reasoning about the message. The two concepts are similar to Kahneman’s distinction about the fast, heuristic-based System I and the slower, deliberative System II [11]. Surface credibility is message credibility based on System I reasoning, while earned credibility is message credibility based on System II reasoning. Research results [28] have established that most users evaluate Web page credibility quickly, in a matter of minutes (three minutes are enough for most Web page credibility evaluations). These results are relevant for our experiment design. In order to begin to understand brain activity during message credibility evaluation, we shall limit message design to a single aspect that can be rapidly evaluated.

### 2.3 Research on Brain Activity Related to Message Credibility Evaluation

To our knowledge, message credibility has not been investigated to date using neuroimaging methods. There are, however, some studies that may be in some way associated with our interests. Research conducted on patients with Alzheimers disease and controls showed engagement of hyperactivity of Brodmann Area 38 in Positron Emission Tomography (PET) experiment discussed in [19]. Similarly BA38 was hyperactive in functional Magnetic Resonance Imaging fMRI studies reported in [8] where prospective customers were obliged to choose between similar goods making economic decisions. The activation of BA38 in language related tasks is also reported in [5]. Language and visual perception associations are postulated in [4] where some meta-analyses on BrainMap were conducted. BA10 and BA47 play important role in decision-making related fMRI studies classified by neural networks in [1]. Another fMRI experiment shows that BA10, BA46 and through connectivity BA38 are involved in high- and low-risk



**Fig. 1.** Typical screen shown to participant during the experiment with the long note in the top and short in the bottom.

decisions [7] as well as BA47 reported to be engaged in lexical decisions concerning nonwords or pseudowords in PET/fMRI studies in [22]. Activation of selected BA related to language, decision making and in some way trustworthiness is often investigated using competition or game element like in [21], but in our experiments one should remember there is no game component.

### 3 Experiment Design

We have designed and conducted a pilot experiment to study message credibility evaluations using EEG. The pilot experiment was carried out at Marie Curie Skłodowska University in Lublin, Poland, from June, 15th till July, 14th, 2019 (MCSU Bioethical Commission permission 13.06.2019).

The goal of the pilot experiment was to observe the electrical activity and the most active areas of the participant's brain cortex during tasks involving message credibility evaluation, as well as the influence of the message design on this process. In order to ensure that participants could only rely on message design during the experiment, the experiment was designed so that the participants would not be familiar with the topic of the messages. The chosen topic of the messages concerned the meaning of Japanese kanji signs.

The experiment was designed to create a condition when participants assess truth or falsehood with practically no knowledge of the message subject. Participants were completely unsure about the correct answer. This situation resembles the case when a person who has no knowledge of the subject receives fake news.

The participants to the experiment were right-handed male students without any knowledge of Japanese. A total of 62 participants took part in the pilot experiment.

#### 3.1 Message Credibility Evaluation Task

Participants were requested to choose the meaning of a Japanese Kanji sign, based on a provided explanations (the message). There were 128 different Kanji signs to be assessed and their meaning was described by a single word or by a

longer description (consisting of at most 20 words) in participant’s native language (see Figure 1). The longer descriptions were designed to logically explain the relationship between the shape and meaning of the Kanji sign, for example: “gutter, curves symbolize pipes” In the remainder of this article, we shall refer to the single word and longer descriptions as “short note” and “long note”, respectively.

In half (64) of the cases, the true meaning was given by a single word, and in the remaining 64 cases the true meaning was described by a longer note. In half of the screens the long note was at the top of the screen and the single-word note was at the bottom, while in the other half, the notes were placed in the opposite way.

The participants’ task was formulated in the form of a question: “Does the Japanese character ... mean: ...” (see Figure 1). The meaning contained in the question was always the meaning that was displayed at the top. The participants could answer the question by selecting “Yes” or “No”. The actual answer was not important for the analysis; what actually mattered was whether the answer matched the single-word note or the longer note (it could only be one of the two). The agreement of the answer with the note is equivalent to a positive evaluation of the note’s message credibility.

Such a setup allowed us to register EEG measurements in four cases:

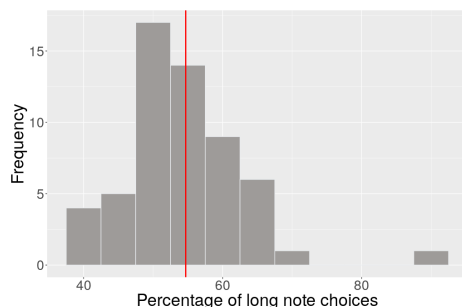
1. ST-SC: Short Top-Short Chosen. The short note was on top and the long note was at the bottom. The short note was chosen as the proper meaning of the kanji sign,
2. ST-LC: Short Top-Long Chosen. The short note was on top and the long note was at the bottom. The long note was chosen as the proper meaning of the kanji sign,
3. SB-SC: Short Bottom-Short Chosen. The short note was at the bottom and the long note was on top. The short note was chosen as the proper meaning of the kanji sign,
4. SB-LC: Short Bottom-Long Chosen. The short note was at the bottom and the long note was on top. The long note was chosen as the proper meaning of the kanji sign.

This also could indicate whether participants prefer the length of the note or their position in the interpretation of Kanji signs.

### 3.2 EEG measurements

In the empirical experiments we were using top EEG devices. We were equipped with a dense array amplifier recording cortical activity with up to 500 Hz frequency through 256 channels HydroCel GSN 130 Geodesic Sensor Nets provided by Electrical Geodesic Systems (EGI). In addition, in the EEG Laboratory the Geodesic Photogrammetry System (GPS) was used. The position of the electrodes is precisely defined in EGI Sensor Nets documentation.

The responses of the cohort of 62 participants were examined. Showing the screen with two notes (single-word and long) was treated as the stimulus (event)



**Fig. 2.** Histogram of frequency of evaluating long note as credible among all participants.

evoking the ERP. For each of 256 electrodes the mean electric charge was calculated in the way as follows: first the ERP was estimated in the interval of 900 ms (beginning 100 ms before showing the stimulus and ending 800 ms after showing that), next the source localisation algorithm sLORETA (GeoSource Parameters set as follow: Dipole Set: 2 mm Atlas Man, Dense: 2447 dipoles Source Montages: BAs) was applied to each signal from each electrode and the average electric current varying in time and flowing through each BA was tabularized. Finally, the Mean Electric Charge (MEC) flowing through each BA was calculated by integration of electric current in 10 ms intervals of time. The procedure of calculating MEC has been described in detail in [31] and using the MEC as one more method of quantitative EEG analysis has been verified in different ways and discussed in [30][32].

### 3.3 Experiment hypotheses

We formulated the following hypotheses:

1. The length of the note has a significant positive influence on the participant's decision about message credibility.
2. The length of the note has a significant influence on Brodmann areas' activity during making decisions about message credibility.
3. The decision of participants about message credibility can be predicted based on measurements of mean electric charges in participant's brains.
4. There are significant differences in the models that predict decisions of participants who frequently choose the long note as compared to models of other participants.

Hypotheses 1 is not directly related to participants' brain activities. Rather, it is a test of our experiment's internal validity. A positive validation of hypothesis 1 would confirm that there exists a relationship between the main independent variable of our experiment and the participant's decision. Such a relationship would confirm the internal validity of our experiment. Hypothesis 2 is related to the first and second research questions. To validate this hypotheses, we need to

BA	Anatomical structure	Known functions
L-BA19	Associative Visual Cortex (V3)	processing phonological properties of words, confrontation naming
L-BA38	temporopolar area	semantic processing, naming of items learned in early life, word retrieval for specific entities, lexico-semantic ambiguity processing, narrative comprehension
L-BA41	Primary Auditory Cortex	working memory
L-BA34	entorhinal area, superior temporal gyrus	working memory, memory encoding, experiencing/processing emotions
R-BA38	temporopolar area	irony processing
L-BA40	Supramarginal Gyrus	Semantic processing, verbal creativity, deductive reasoning
R-BA39	Somatosensory Association Cortex	language processing, literal sentence comprehension, word comprehension (imageability)

**Table 1.** The Brodmann Areas manifested statistically significant differences in MEC measured during choice of long or short note during the experiment, and were selected by the model trained on experimental data to predict long or short note choices. Presented with their corresponding anatomical structure of the brain and known functions as listed in [17]. The prefix of particular BAs stands for L-left and R-right hemispheres.

study brain activity during the experiment and compare this activity in cases when the short note is evaluated as credible (ST-SC or SB-SC) and when the long note is chosen (ST-LC or SB-LC).

Hypothesis 3 can be validated by constructing a classifier that will predict the (binary) decision of participants with sufficiently high accuracy. However, the validation of 4 requires training two such classifiers, one based on the set of participants who tend to evaluate long messages as credible, and another on the remaining set of participants. The comparison of these two classifiers is only possible if the two classifiers are explainable, which excludes the use of black-box classifiers such as neural networks.

## 4 Experiment Results

After the recording phase of the experiment the EEG signal of 57 participants was analysed. Each participant answered 128 questions requiring making decision by choosing the short or the long note as a correct answer. That gave  $57 \times 128 = 7296$  responses. In fact, 7296 responses were collected.

The long note was chosen 1782 times when it was on top of the screen (SB-LC), while the short one was chosen 1442 times when it was on top (ST-SC). On the other hand, the long note was chosen 2206 times when it was at the bottom (ST-LC), as compared to 1866 times when the short note was chosen while at the bottom (SB-SC). This means that no matter whether the long note was on top or at the bottom of the screen it was chosen more frequently and the total difference is 680 (or 9.3%) towards the long note. The length of the note influences the participants' decision regardless of the note's position. This positively verifies hypothesis 1.

We can also test whether or not the message credibility evaluation was random. Participants could evaluate the long or short note as credible. If the choice is be random, the choices of the long or short note should form a binomial dis-



No.	BAs used by the model	Accuracy.	Precision	Recall	F1 Score
1	LBA19, LBA38, LBA41,LBA34, RBA38 ,LBA40,RBA39	0.793	0.786	0.786	0.786
2	LBA19,RBA13,LBA38,LBA41,LBA34, ,RBA38,LBA40,RBA39	0.793	0.714	0.833	0.769
3	LBA13,LBA19,RBA13,LBA38,LBA41 , LBA34,RBA38,LBA40,RBA39	0.793	0.786	0.786	0.786
4	LBA13,LBA19,LBA38,LBA41,RBA41 , LBA34,RBA38,LBA40,RBA39	0.793	0.714	0.833	0.769
5	LBA13,LBA34,RBA38,RBA47,RBA39	0.793	0.786	0.786	0.786

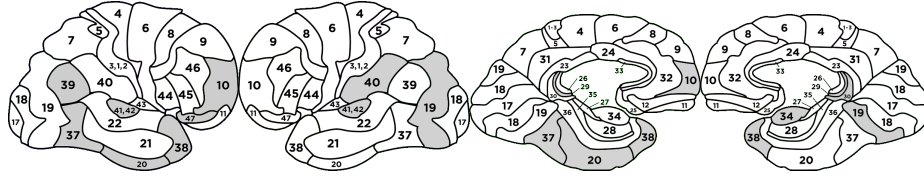
**Table 2.** Five best models found by the brute force method to classify participants’ decisions using selected BAs from the set of BAs manifesting statistically significant differences in activity. For more details see text.

tribution with probability 0.5. We used the binomial test and calculated the p-value, which was less than 0.00001 (we observed 3988 choices of the long note out of 7296 message credibility evaluations). Therefore, we concluded that we could reject the possibility that the choices of notes in the experiment were binomially random.

We also found that a subset of participants who chose the long note more frequently than other participants. For each participant, we calculated how frequently they evaluated the longer note as credible. This frequency is shown on a histogram in Figure 2. The median of this distribution, which is shown as red vertical line in Figure 2, is approximately 54% because overall the long note was evaluated as credible more frequently than the single word note. 25 of 57 participants chose the long note in more than 55% of all possible choices, and 17 participants evaluated the long note as credible in more than 60% of cases.

Statistically significant differences in MEC of all participants were observed using the Mann-Whitney-Wilcoxon test for choosing long/short note in the time between 150 ms to 650 ms from stimulus, but only in the case when the short note was on top of the screen. The spontaneous decisions in ERP experiments are usually made in such interval [32][30][31][23][16]. If the short note was at the bottom, there were no statistical differences in MEC observed in the cognitive processing time interval. This partially verified the hypothesis 2. Note that in every scenario we asked the participant about the meaning of the note that was explained at top of the screen. This questions most likely affected participants’ focus. For this reason, in the analysis we shall focus on the comparison of the cases ST-SC and ST-LC.

The BAs for which there was a statistically significant difference in Mann-Whitney-Wilcoxon Test in MEC during the interval [150ms, 650ms] from the stimulus are: L-BA19, L-BA38, L-BA41, L-BA34, R-BA38, L-BA40, R-BA39, R-BA10, L-BA13, R-BA13, R-BA20, L-BA29, R-BA29, L-BA30, R-BA37, R-BA41, L-BA42 and R-BA47. However, only a subset of them was chosen by the model of participants credibility evaluations of the long or short notes, as described in the next section. This subset of 7 most significant BAs is shown on Table 1.



**Fig. 3.** Brodmann areas most significant for predicting message credibility evaluations

#### 4.1 Regression Model of Message Credibility

We used a generalized logistic regression classifier to predict message credibility evaluations. The classifier was trained on evaluations of a subset of 42 (75%) participants and validated on evaluations of 15 participants. This means that the classifier had 42 observations of each class available for training. Only the BAs manifesting statistically significant differences in their activity were taken into consideration as independent variables. These were BA13, BA19, BA29, BA30, BA34, BA38, BA40, BA41, BA42 in the left hemisphere and BA10, BA13, BA20, BA29, BA37, BA38, BA39, BA41, BA47 in the right hemisphere see. Table 1.

A set of independent variables for the logistic regression classifier was constructed in a special manner, using the following reasoning:

1. 18 different BAs had significant differences during the experiment.
2. It is not obvious that all of them should be taken as features of classifiers.
3. If so, then each subset of these 18 BAs could be considered as independent variables for classifiers.
4. Why not to check all possible combinations of all subsets and then find the ideal model for the whole population?

There are 262143 statistically significant BAs possible subsets as a sum of combinations as in the following equation:  $N = \sum_{i=1}^Q \binom{Q}{i}$  where  $Q = 18$ .

The Logistic-regression classifiers were built for all subsets in the above-mentioned manner. Results achieved by 5 best classifiers are presented in Table 2. Note that there are only 7 BAs engaged in the best classifier and also used in the following 3 next-best classifiers: LBA19, LBA38, LBA41, LBA34, RBA38, LBA40, RBA39. This confirms that the proposed methodology of selecting variables produces models that have consistently similar sets of independent variables. The 7 BAs used in the best classifier are clearly the most useful independent variables overall.

For the best combination of independent variables, we have used 10-fold cross-validation to confirm our results. The average efficiency achieved by classifiers in this cross-validation is: Accuracy 0.68, Precision 0.6582, Recall 0.7184 and F1 Score 0.7184 which proves its stability.

Such classification results positively confirm the hypothesis 3.

#### 4.2 Modeling brain activity of participants that prefer long notes

The simulations described above were repeated in order to find the best classification models for the participants who found long notes credible more frequently.

The goal was to compare the best classifiers for this subset of participants (25 out of 57, or 43% of all participants who evaluated the long note as credible in more than 55% of the cases) to the best classifiers for the remaining participants. Note that the classifier for participants who preferred long notes had 25 observations of each class (single-word, or long note selection) in the training set and in the validation set.

For participants who were choosing long notes more frequently, the following BAs manifest statistically significant differences in activity: LBA13, RBA25, LBA19, LBA30, RBA29, LBA29, LBA46, RBA13, RBA19, RBA37, LBA34, RBA47, RBA39. The best classifier for the long-note-choosers involved: LBA19, RBA29, LBA29, LBA46 and reached the characteristics of accuracy: 0.923, precision: 0.833 and recall: 1.000. For the best combination of independent variables, we have used 10-fold cross-validation to confirm our results. The average efficiency achieved by classifiers in this cross-validation is: Accuracy 0.6615, Precision 0.6285, Recall 0.6562 and F1 Score 0.909 which also proves its stability. The BAs that were most significant for classification of participant's decisions are visualized on Figure 3.

The BA29 (granular retrosplenial cortex) associated with memory retrieval and emotion related to language appeared in the best models of participants who preferred long notes in contrast to the set of best models generated for the entire population. This positively verified hypothesis 4, as it was a significant difference in the models of message credibility decision for the two groups.

### 4.3 Discussion and Limitations

The aim of this article was to address three research questions.

Firstly, we wondered if it is possible to indicate the brain regions that are engaged in receiver's cortex while evaluating the message credibility? Indeed, we found 18 BAs that manifested statistically significant differences during tasks of message credibility evaluation.

Some of the BAs: BA38, BA10, BA46, BA47 confirmed previous findings from the experiments where decision-making was involved. Our models, however, chose the best candidates for future research of brain functional anatomy related to credibility evaluation.

It was also interesting to check whether the design of the message had the influence on brain activity? The experiment was designed so that participants had no knowledge of the message subject and could only be influenced by structure of the note that described the Kanji sign. Participants had a tendency to choose a longer note, and we found statistically significant differences in brain activity when a longer note was chosen, as compared to the situation of choosing a single-word note.

The most interesting results were observed while addressing the third research question, i.e. whether it is possible to model and predict brains' message credibility evaluation?

Only 18 BAs that were statistically significantly different in the spectrum of MEC were taken into consideration, and, therefore, it was possible to find the

best classifier for the entire investigated population. The chosen method, consisting of checking all possible combinations of BA subsets, was computationally demanding and required parallel computations.

The analysis of independent variables used by the best classifiers of message credibility revealed a set of 7 BAs that are the best independent variables to predict message credibility evaluations based on EEG measurements. The classifiers achieve a satisfactory accuracy of 79% (against the baseline of 50%). This result is a step forward in the study of brain functional anatomy, as this is a first finding concerning areas in the brain involved in message credibility evaluation. It is also an achievement in the field of neuroinformatics, as our results give a firm basis for the development of an EEG-based measurement method of message credibility. This method could have many practical applications, as for example in evaluating messages aimed at combating and debunking fake news. Note that our models used generalized logistic regression, as our goal was to have an explanatory model. The use of other classifiers in the future could most likely further improve the accuracy of predicting message credibility evaluation.

We attempted to investigate incorrect message credibility evaluations as well. In our experiment, participants had no access to ground truth. However, the experiment was designed to create a situation when participants could choose to be influenced by message design in their credibility evaluation. The only factor of message design used in the experiment was message length. We found that a subset of participants tended to more frequently evaluate long notes as credible, as compared to other participants. By modeling the decisions of this subset of participants separately, we identified the area in the brain (BA29, the granular retrosplenial cortex) associated with memory retrieval and emotion related to language that has impact on message credibility evaluations of participants who preferred long notes. While this finding needs to be verified on larger samples, it demonstrates the possibility of studying other factors that impact message credibility evaluation using EEG measurements.

## 5 Conclusion and Future Work

Although the subject of credibility has received significant attention from psychologists, media scientists, and computer scientists aiming at detecting and combating fake news, it had not been studied before using brain activity measurements. This is probably due to the fact that it is difficult to design and carry out brain measurement experimentation. As a result, no method of message credibility measurement (or any credibility measurement) based on brain activity had been proposed before.

We have described in this article the first experiment designed to study message credibility evaluation. The experiment controlled the influence of participants' knowledge and beliefs on message credibility evaluation by using a subject with which the participants were not familiar (Kanji signs). Furthermore, the experiment controlled message design factors that could influence message credibility evaluation and limited them to one factor, i.e. message length.

We have identified and described 7 BAs that have the greatest impact on message credibility evaluations. The selection of these brain functional areas was

done by a statistical analysis of EEG signals and based on generalized logistic regression classifiers of participants' decisions. These classifiers achieved an accuracy of 79%, demonstrating the feasibility of creating an EEG-based message credibility measurement method.

Our experiment has several limitations due to its design. One limitation is the small amount of users who are influenced by the message design. We intend to continue investigating the impact of message design using other design factors, such as message persuasiveness.

In the future, we plan to run similar experiments that require basic knowledge of Japanese by participants, in order to observe the impact of participant's own knowledge on message credibility evaluation.

## References

1. Ahmad, F., Ahmad, I., Dar, W.M.: Identification and classification of voxels of human brain for reward-related decision making using ann technique. *Neural Computing and Applications* **28**(1), 1035–1041 (2017)
2. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. *Journal of economic perspectives* **31**(2), 211–36 (2017)
3. Appelman, A., Sundar, S.S.: Measuring message credibility: Construction and validation of an exclusive scale. *Journ. & Mass Comm. Quart.* **93**(1), 59–79 (2016)
4. Ardila, A., Bernal, B., Rosselli, M.: Language and visual perception associations: meta-analytic connectivity modeling of brodmann area 37. *Behavioural neurology* (2015)
5. Ardila, A., Bernal, B., Rosselli, M.: The elusive role of the left temporal pole (ba38) in language: a preliminary meta-analytic connectivity study. *International Journal of Brain Science* **2014** (2014)
6. Borah, P., Xiao, X.: The importance of 'likes': The interplay of message framing, source, and social endorsement on credibility perceptions of health information on facebook. *Journal of health communication* **23**(4), 399–411 (2018)
7. Cohen, M., Heller, A., Ranganath, C.: Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Cognitive Brain Research* **23**(1), 61–70 (2005)
8. Deppe, M.e.a.: Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making. *Journal of Neuroimaging* **15**(2), 171–182 (2005)
9. Forgas, J.P., Baumeister, R.: *The Social Psychology of Gullibility: Conspiracy Theories, Fake News and Irrational Beliefs*. Routledge (2019)
10. Hovland, C.I., Weiss, W.: The influence of source credibility on communication effectiveness. *Public opinion quarterly* **15**(4), 635–650 (1951)
11. Kahneman, D.: *Thinking, fast and slow*. Macmillan (2011)
12. Kakol, M., Jankowski-Lorek, M., Abramczuk, K., Wierzbicki, A., Catasta, M.: On the subjectivity and bias of web content credibility evaluations. In: *Proceedings of the 22nd international conference on world wide web*. pp. 1131–1136. ACM (2013)
13. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., et al.: The science of fake news. *Science* **359**(6380), 1094–1096 (2018)
14. Liu, X., Nielek, R., Adamska, P., Wierzbicki, A., Aberer, K.: Towards a highly effective and robust web credibility evaluation system. *Decision Support Systems* **79**, 99–108 (2015)

15. Meijer, E.H., Verschuere, B.: Deception detection based on neuroimaging: Better than the polygraph? *Journal of Forensic Radiology and Imaging* **8**, 17–21 (2017)
16. Mikołajewska, E., Mikołajewski, D.: The prospects of brain—computer interface applications in children. *Open Medicine* **9**(1), 74–79 (2014)
17. Pascual-Leone, A., Davey, N.J., Rothwell, J., Wasserman, E.M., Puri, B.K.: *Handbook of transcranial magnetic stimulation*, vol. 15. Arnold London (2002)
18. Rafalak, M., Abramczuk, K., Wierzbicki, A.: Incredible: Is (almost) all web content trustworthy? analysis of psychological factors related to website credibility evaluation. In: *Proceedings of the 23rd international conference on world wide web*. pp. 1117–1122. ACM (2014)
19. Rinne, J.e.a.: Semantic decision making in early probable ad: a pet activation study. *Cognitive brain research* **18**(1), 89–96 (2003)
20. Rutjens, B.D., Brandt, M.D.: *Belief systems and the perception of reality*. Routledge (2018)
21. Sailer, U., Robinson, S., Fischmeister, F.P.S., Moser, E., Kryspin-Exner, I., Bauer, H.: Imaging the changing role of feedback during learning in decision-making. *Neuroimage* **37**(4), 1474–1486 (2007)
22. Specht, K., Holtel, C., Zahn, R., Herzog, H., Krause, B.J., Mottaghy, F.M., Radermacher, I., Schmidt, D., Tellmann, L., Weis, S., et al.: Lexical decision of nonwords and pseudowords in humans: a positron emission tomography study. *Neuroscience Letters* **345**(3), 177–181 (2003)
23. Tadeusiewicz, R., et al.: *Neurocybernetyka teoretyczna*. Wydawnictwa Uniwersytetu Warszawskiego (2009)
24. Tseng, S., Fogg, B.: Credibility and computing technology. *Communications of the ACM* **42**(5), 39–44 (1999)
25. Turek, P., Wierzbicki, A., Nielek, R., Datta, A.: Wikiteams: How do they achieve success? *IEEE Potentials* **30**(5), 15–20 (2011)
26. Wang, H., Chang, W., Zhang, C.: Functional brain network and multichannel analysis for the p300-based brain computer interface system of lying detection. *Expert Systems with Applications* **53**, 117–128 (2016)
27. Wawer, A., Nielek, R., Wierzbicki, A.: Predicting webpage credibility using linguistic features. In: *Proceedings of the 23rd international conference on world wide web*. pp. 1135–1140. ACM (2014)
28. Wierzbicki, A.: *Web Content Credibility*. Springer (2018)
29. Wierzbicki, A., Turek, P., Nielek, R.: Learning about team collaboration from wikipedia edit history. In: *Proceedings of the 6th international symposium on Wikis and open collaboration*. pp. 1–2 (2010)
30. Wojcik, G.M., Masiak, J., Kawiak, A., Kwasniewicz, L., Schneider, P., Polak, N., Gajos-Balinska, A.: Mapping the human brain in frequency band analysis of brain cortex electroencephalographic activity for selected psychiatric disorders. *Frontiers in neuroinformatics* **12** (2018)
31. Wojcik, G.M., Masiak, J., Kawiak, A., Schneider, P., Kwasniewicz, L., Polak, N., Gajos-Balinska, A.: New protocol for quantitative analysis of brain cortex electroencephalographic activity in patients with psychiatric disorders. *Frontiers in neuroinformatics* **12** (2018)
32. Wojcik, G.M., Masiak, J., Kawiak, A.T., Kwasniewicz, L.K., Schneider, P., Postepski, F., Gajos-Balinska, A.: Analysis of decision-making process using methods of quantitative electroencephalography and machine learning tools. *Frontiers in Neuroinformatics* **13**, 73 (2019)