HAL
open science

Towards Implicit Visual Memory-Based Authentication

Claude Castelluccia, Markus Duermuth, Maximilian Golla, Fatma Deniz

To cite this version:

Claude Castelluccia, Markus Duermuth, Maximilian Golla, Fatma Deniz. Towards Implicit Visual Memory-Based Authentication. Network and Distributed System Security Symposium (NDSS), ISOC, Feb 2017, San Diego, United States. hal-01109765

HAL Id: hal-01109765

https://inria.hal.science/hal-01109765v1

Submitted on 11 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards Implicit Visual Memory-Based Authentication

Claude Castelluccia
Inria Grenoble
claude.castelluccia@inria.fr

Markus Dürmuth and Maximilian Golla
Ruhr-University Bochum
{markus.duermuth,maximilian.golla}@rub.de

Fatma Deniz
University of California, Berkeley
fatma@berkeley.edu

***Abstract*—Selecting and remembering secure passwords puts a high cognitive burden on the user, which has adverse effects on usability and security. Authentication schemes based on implicit memory can relieve the user of the burden of actively remembering a secure password. In this paper, we propose a new authentication scheme (MooneyAuth) that relies on implicitly remembering the content of previously seen Mooney images. These images are thresholded two-tone images derived from images containing single objects. Our scheme has two phases: In the *enrollment phase*, a user is presented with Mooney images, their corresponding original images, and labels. This creates an implicit link between the Mooney image and the object in the user's memory that serves as the authentication secret. In the *authentication phase*, the user has to label a set of Mooney images, a task that gets performed with substantially fewer mistakes if the images have been seen in the enrollment phase. We applied an information-theoretical approach to compute the eligibility of the user, based on which images were labeled correctly. This new *dynamic scoring* is substantially better than previously proposed static scoring by considering the *surprisal* of the observed events. We built a prototype and performed three experiments with 230 and 70 participants over the course of 264 and 21 days, respectively. We show that MooneyAuth outperforms current implicit memory-based schemes, and demonstrates a promising new approach for fallback authentication procedures on the Web.**

I. Introduction

User authentication is an essential requirement for modern websites as more and more access-controlled services move online. Existing user authentication schemes are commonly based on *something you know*, such as passwords, *something you have*, such as secure tokens, or *something you are*, such as biometry. These authentication schemes suffer from the competing requirements of security and usability [8], which are hard to fulfill simultaneously. Furthermore, users seem to disfavor password-based authentication [32], [23], hence alternative schemes are becoming necessary [1]. Despite substantial research effort to improve the state-of-the-art, currently deployed authentication methods are far from optimal.

Permission to freely reproduce all or part of this paper for noncommercial purposes is granted provided that copies bear this notice and the full citation on the first page. Reproduction for commercial purposes is strictly prohibited without the prior written consent of the Internet Society, the first-named author (for reproduction of an entire paper only), and the author's employer if the paper was prepared within the scope of employment.
NDSS '17, 26 February - 1 March 2017, San Diego, CA, USA
Copyright 2017 Internet Society, ISBN 1-1891562-46-0
http://dx.doi.org/10.14722/ndss.2017.23292

This paper explores a new type of knowledge-based authentication scheme that eases the high cognitive load of explicit passwords and thus has the potential to improve the usability and security of knowledge-based authentication. In particular, we study how *implicit memory* can be used to design a usable, deployable, and secure authentication scheme.

Current knowledge-based authentication schemes are based on *explicit memory*, where users are asked to create a random combination of characters as their authentication secret and to explicitly provide this secret at the time of authentication. Such secrets are usually very difficult to remember as one has to work consciously to remember the specific secret. In contrast, with an implicit memory-based scheme, users first learn an association between a task and its solution. This learned association is then used as the authentication secret. Because recalling a situation that is stored in the implicit memory is remembered with less effort [36], [37], almost unconsciously, such an authentication scheme relieves users of the high cognitive burden of remembering an explicit password. This has the potential to offer usable, deployable, and secure user authentication.

In this work, we built a novel, operational authentication scheme utilizing implicit memory based on Mooney images [31]. A Mooney image is a degraded two-tone image of a single object. This object is usually hard to recognize at first sight and becomes easier to recognize when the original image was presented to the user.

Our scheme is composed of two phases: In the *enrollment phase*, the user learns the association between a set of Mooney images, their original versions, and labels describing the content of the image. This process is also called "priming". During the *authentication phase*, a larger set of Mooney images, including the primed Mooney images from the enrollment phase, are displayed to the user. The user is then asked to provide a label for the hidden object in each Mooney image. Using our *dynamic scoring* algorithm, the system computes an authentication score and provides or denies access accordingly. Due to relatively slow enrollment and authentication, the current scheme seems particularly suited for fallback authentication, also known as account recovery.

We conducted three experiments to identify practical parameters, to measure long-term effects, and to determine the performance of the scheme. We conducted Experiment 1 over the course of 25 days with 360 participants of which 230 finished both phases. The results of this experiment were used for parameter selection. To identify long-term priming effects

of Mooney images we re-invited the participants after 264 days in Experiment 2. To validate the overall performance of the scheme, we performed Experiment 3 with 70 new participants over the course of 21 days.

A. Contributions

Our contributions include:

1) We present a novel authentication scheme, based on implicit visual memory, that outperforms existing ones in terms of false acceptance and false rejection rates, as well as the time required for authentication.
2) To decide whether a user successfully passes the authentication phase, the inputs have to be evaluated, i. e., "scored." We propose an alternative scoring technique, *dynamic scoring*, which is inspired by the notion of *self-information*, also known as *surprisal*. We show that our scoring technique substantially outperforms the static scoring proposed in previous work by Denning et al. [16].
3) We demonstrate the practicability of our scheme by implementing it and conduction three experiments. The results show that MooneyAuth substantially outperforms current implicit memory-based authentication schemes [16].
4) We are the first to study long-term priming effects of Mooney images over a period as long as 8.5 months. The results reveal a substantial long-term priming effect for Mooney images, which implies that MooneyAuth is suited for fallback authentication with long intervals between enrollment and authentication.

B. Related Work

1) Implicit Memory-Based Authentication Schemes: Applying the knowledge about how humans store and recall information was first applied to user authentication by Weinshall and Kirkpatrick [43]. However, the proposed scheme made use of the explicit characterization of images that were stored in human memory, not implicit memory, and the performance of the proposed scheme is unsuitable for deployment.

The scheme that comes closest to MooneyAuth, using implicit memory and the priming effect, was proposed by Denning et al. [16]. They presented an authentication scheme based on implicitly learning associations between degraded drawings of familiar objects (e .g., animals, vehicles, or tools) and their complete drawings. Each degraded drawing was created by using fragmented lines instead of continuous lines. In their paper, the authors presented a preliminary authentication scheme and performed a user study. Their results show that many of these drawings show a (small) priming effect, but this effect is too small for using it in an authentication scheme for all but two images they tested. As agreed by the authors, the viability of such system concept is dependent upon being able to systematically identify or create images with a sufficiently strong priming effect. Our paper builds on this work to propose a complete and efficient system. We show that Mooney images provide a strong priming effect necessary to implement such a practical scheme, and we build a real prototype.

Bojinov et al. also proposed the concept of implicit learning to design a scheme that resists coercion attacks where the user is forcibly asked by an attacker to reveal the key [6]. The proposed scheme is based on a crafted computer game. While the secret can be used for authentication, the participant cannot be forced to reveal it since the user has no conscious knowledge of it. The authors perform a number of user studies using Amazon's Mechanical Turk to validate their scheme. Although the proposed idea is very interesting, performance results show that their scheme is not practical and cannot be used for real-world applications: the registration phase takes more than 45 minutes for a single password and put a lot of cognitive burden on users.

2) Graphical Authentication Schemes: A number of graphical password schemes share the idea of identifying trained images from a set of decoy images. Probably the most well-known scheme is PassFaces [34], which is based on the human capability to recognize familiar faces. During authentication, a set of faces is shown to a user, where the user selects known ones. However, user bias in the selection of the images renders the system vulnerable to guessing attacks [14].

The authentication scheme *Use Your Illusion* [22] also uses degraded versions of images, i. e., blurred images, but still relies on explicit memory. In this scheme, users are required to generate an image portfolio, explicitly learn and memorize the images belonging to the portfolio, i. e., via spaced repetitions, and finally authenticate by re-identifying the set of images that were distorted by an oil-painting filter. In contrast to our evaluation, their user study did not test the long-term performance, so it is unknown how this scheme performs over time and whether it decreases the user's cognitive burden. Furthermore, the security of Use Your Illusion heavily depends on the image degrading algorithm and its parameters, which might be identifiable by a computer vision algorithm. In contrast, we assume all images and their Mooney versions including the labels are known to the attacker and show that even in this case MooneyAuth is secure. Hence, the security of MooneyAuth does not rely on the fact that there is no algorithm, which can recover the original image and its label. Other graphical authentication schemes, based on explicit memory, are surveyed in [4].

3) Fallback Authentication Schemes: One prominent use case for MooneyAuth is fallback authentication. Such schemes are used to help users recover their forgotten passwords. Depending on the deployed system, the effort (authentication time, and workload) can be higher than in primary authentication systems, which are used on a daily basis. However, the authentication secret obviously requires to be memorable for longer timeframes, since it must still be available in the case the primary means of authentication, e. g., the password is forgotten.

The range of currently proposed or deployed fallback authentication systems is not truly satisfying. The most frequent systems are based on password reset via out-of-band communication or personal knowledge questions. In the former case, a registered email address, a mobile phone number, or a mobile app on a smartphone of the user [18] is used to send the original password, a new password, or a time-limited password reset link. However, receiving such password reset messages can be risky if not correctly implemented [10] and can be error-prone if the contact details on record are out of date. Furthermore, not all users like the idea of giving out their cellphone number or email address due to privacy concerns.

Even worse, receiving messages is not always possible, i. e., using mobile data while abroad, or the receiving device is not available.

If communicating over a secondary channel is not possible, personal knowledge questions, sometimes called cognitive passwords, are used. The security of such systems is well studied [45]. However, as demonstrated by a number of recent work [7], [9], [38], [21], their security is rather low, as the secret answers to the asked questions can be easily guessed.

Renaud et al. [35] introduced a scheme based on associations between images and text (selected by the user) as a potential replacement for security questions. However, their approach suffers from the drawbacks of explicit memory-based schemes.

Authentication using information about the social graph of a user, so-called social authentication, has been demonstrated [12], [39]. Facebook deployed such a social scheme called *Trusted Contacts*. As a secondary fallback-mechanism users can choose up to five friends that receive parts of a recovery code via email in the case the user has forgotten the password. By collecting three or more parts of the code one is able to reset the password. However, typical recovery times can quickly rise from hours to days, which is a potential drawback of this approach.

4) Associative and Repetitive Memory-Based Authentication Schemes: Recent work by Bonneau and Schechter [11] demonstrated that users are capable of remembering cryptographically-strong secrets via spaced repetition. In their experiment, they enabled users to learn a limited number of strong authentication secrets by displaying an additional code that was required to login. This code did not change and was only shown after an annoying delay which was increased at every login attempt. The users were motivated to accelerate the login procedure and not wait for the code to display, by entering the code, which they subliminally learned by heart, due to its continuous repetition. After some of such fast and successful logins, the code was extended.

A similar user study, realized by Blocki et al. [5], improved the repetition idea. Based on so-called Person-Action-Object (PAO) stories they were able to combine associative and repetitive memory to improve the concept. They asked their participants to invent a story based on a shown photo, a user-chosen famous person, and a randomly selected action-object pair that served as authentication secret. In contrast to [11], the users were able to see the complete secret at once and were told that they are required to learn the secret. Finally, the users were able to remember their secrets for longer times with fewer rehearsals due to the PAO story mnemonic.

C. Outline

This paper is structured as follows: Section II introduces the concept of implicit memory and Mooney images. Our scheme is described in Section III. Then, we present details on the three experiments we performed. First, the pre-study for estimating the required parameters in Section IV, a long-term study proving that the priming effect of Mooney images last over time in Section V, and the main study demonstrating the general performance of the scheme in Section VI. We discuss security properties in Section VII, and conclude with some final remarks in Section VIII.

II. Background

A. Explicit vs. Implicit Memory

Explicit memory is a type of memory that is based on intentional recollection of information with the purpose to consciously recall this information at a later time. We use this type of memory, also referred to as *declarative memory*, constantly in our daily life [20]. For example when we remember the time of our flight the next day, recall our address, or a chain of strings that forms our passwords.

In contrast, *implicit memory* relies on the unintentional recollection of information. In this case, we are not aware of the specific information we stored in our memory, but we can easily recall the information. This type of memory, also referred to as *nondeclarative memory*, can usually be observed in habitual behavior, such as riding a bicycle or playing an instrument [20]. The cognitive and neural mechanisms of explicit and implicit memory are not entirely understood [19]. Some studies suggest a distinct mechanism for explicit and implicit memory [36], [30], whereas others suggest a joint mechanism [3], [42]. One way to trigger implicit memory is an effect called priming [29], [13]. Priming occurs when the previous exposure (conscious or unconscious) to a stimulus affects the performance of a subsequent task. For example, when a series of images with specific objects (primes) are presented to the participants, their recognition performance (e. g., time and correctness) of a similar object in another or the same image that is presented later improves. Throughout this paper, we use such priming effects that are based on repetition and association. In a first enrollment phase, we present participants an association between a thresholded Mooney image and the original image with a label. In a second authentication phase, we repeat the previously primed Mooney image (among other non-primed Mooney images) and measure the recognition performance of the repeated image. In some cases, priming has been shown to have long-lasting effects [13].

B. Mooney Images

A Mooney image is a thresholded, two-tone image showing a single object. This object is hard to recognize at first sight with recognition times in second to minute range [25]. In some cases, the recognition is abrupt and gives rise to a feeling of having solved a difficult problem (also known as the aha-feeling or Eureka-effect) [27]. This abrupt recognition can happen intrinsically [25], after the contour of the object is marked [41], or after presenting the subject with the original image [28], [24], [17]. Once a subject has seen the original grayscale image from which the Mooney is generated, recognition is much accelerated. An example of a Mooney image is presented in Figure 1[1].

The value of using Mooney images for authentication is that they are very likely to trigger brain processes that are involved in implicit memory [2]. Implicit memory, as

[1] To understand the effect of Mooney images, we suggest the reader to spend some time trying to identify the object in Figure 1, and then look at Figure 7 at the end of this paper.

stated above, does not require direct conscious involvement but happens with less effort in comparison to explicit memory. Triggering the implicit memory for authentication is therefore desirable as it reduces the cognitive load for users. Priming is one way to trigger implicit memory and Mooney images are excellent example that can be used to prime participants to specific concepts.

Fig. 1. Example of a Mooney image. To understand the effect of Mooney images, we suggest to spend some time trying to identify the hidden object, and then to look at Figure 7 at the end of this paper.

III. The MooneyAuth Scheme

In the next section, we describe the basic construction of our authentication scheme. We first describe how Mooney images are generated, and then present the two phases, enrollment, and authentication, of our protocol.

A. Mooney Image Generation

In this work, we use an extended set of two-tone, Mooney images that contain not only faces as used originally [31], but also objects (e. g., animals, fruits, or tools) of different types [17], [25].

We selected our Mooney images from an automatically generated two-tone, Mooney image database [25]. This database is based on a large number of images collected from the Web. First, concrete nouns were selected from a linguistic database [44] (based on the directness of reference to sense experience, and capacity to arouse nonverbal images, cf. [25]). These words were used as search terms to automatically download images from an online image database. Second, the images were converted to grayscale and were smoothed using a 2D smooth operation with a Gaussian kernel (σ = 2 pixels and full width at half maximum (FWHM) = 5 pixels). Third, images were resized to have a size of 350×350 pixels (subsampled with an appropriate scale factor). These parameters were selected to create Mooney images that are hard to recognize by a user at first sight [26]. The smoothing operation is in particular important for the results as the thresholding algorithm applied in the next stage operates better on smoothed images than on not smoothed ones. Lastly, the smoothed and resized images were thresholded using a histogram based thresholding algorithm (Otsu's thresholding method [33]) to generate the Mooney images. This thresholding method assumes that each image has two classes of pixel properties: A foreground and a background. For each possible threshold, the algorithm iteratively computes the separability of the two classes and converges when the maximum separability is reached. Once the images are automatically downloaded and thresholded, a manual clean up session by human subjects needs to be done. This manual cleaning session is necessary because some images that are automatically downloaded from the web may not include the object that corresponds to the search word (e. g., cat), hence, need to be removed from the image set [15]. Subsequently, a selection of suitable Mooney images took place. While the original Mooney image database contained 330 images [25], for our experiments we considered images with a mean recognition rate of 5 seconds and longer resulting in 250 images. We further reduced this set to 120 images to obtain enough samples per image for an estimated 100 participants.

A suitable Mooney image for the purpose of this application is an image that is difficult to recognize without a previous explicit presentation of the original image. At the same time, if the user has seen the original image then the user should be able to correctly identify and label the hidden object. This procedure makes use of implicit memory as the users first learn the association between the original image and the corresponding Mooney image without an explicit effort. As in the example of riding a bike, users usually do not remember the details of the original image but can name the hidden object in the Mooney image when they have previously seen the original image. For some images, the object shown in the Mooney image can be recognizable by a non-primed user as well, but only after a relatively long time, whereas primed users will recognize it almost instantly. Therefore, within this work, we will treat images with a recognition time beyond a set threshold as "likely not primed".

B. Description

We use a (large) set of images I and their corresponding Mooney images.

Enrollment (Priming) Phase

- When a new user is enrolled, the server first assigns two disjoint subsets $I_P, I_N \subset I$ with $|I_P| = |I_N| = k$ to the user. I_P reflects the primed images, I_N the non-primed images.
- The subset I_P is then used to prime the user. During this session, first a Mooney image, then the original image and a label that describes the object in the image is presented to the user. This procedure creates an association between the original image, the correct label of the image and the corresponding Mooney image.

Authentication (Recall) Phase

- At the beginning of the authentication phase, the two subsets I_P, I_N for this user are retrieved from the database. The primed and non-primed Mooney images ($I_P \cup I_N$) are then presented to the user in a pseudo-randomized order. For each Mooney image presentation, the user is requested to type in the label of the object that the image contains, or skip the image if the user is not able to recognize any object.
- Two metrics are then computed for each image:
  (i) The *correctness of the label* is computed by comparing the typed label to a list of previously defined labels. This

is achieved by a distance metric that measures how similar the label provided by the user matched the defined labels. (ii) The *recognition time*, i. e., the time between displaying the image and the first keystroke. If the recognition time is longer than 20 seconds, we treat the image as if the label were incorrect, i.e., "likely non-primed". (We chose 20 seconds as threshold as we expect the recognition for primed images to occur almost instantaneous, but then allow the user to hesitate a couple of seconds before starting to type the label. From our experience, recognition without being primed takes closer towards a minute to happen.)

- Authentication is based on the hypothesis that the user labels the primed images more often (and faster) correctly than those Mooney images that the user was not primed on. Sometimes primed images will be labeled incorrectly and vice versa. To tolerate some of these errors, we compute a score from the correct and incorrect labels and accept a user if the reached score is above a specific threshold. There are several possibilities to perform this *scoring*. After the necessary terminology is introduced in the next section, we will discuss two scoring methods.

C. Terminology

This section introduces some of the notations that are used throughout this paper. For one specific image with index i (which is displayed to the user) there are four possible events that we need to consider: the image was/was not primed for the user (i. e., it is in I_P or in I_N), and the user provides a correct or an incorrect label for the image. We denote the probability that a (randomly chosen) user correctly labels a primed image with p_i, and the probability that a user correctly labels a non-primed image with n_i. We expect p_i to be larger than n_i, and we denote the difference with $d_i := p_i - n_i$. A positive d_i indicates that priming is working for this image. For a reasonably well-working priming, images should have $d_i > 0.5$. (Those are called "ideal" in [16], which is slightly misleading as "ideal" in a strict sense is $d_i = 1$.) In Section IV we will see that $1/3$ of the total images have $d_i > 0.5$, i. e., we can identify a good amount of images that work well for our authentication scheme.

D. Adversary Model

We consider a strong adversary that has detailed information about the image database I, but has no information about the subsets I_P and I_N.

1) We assume the adversary knows the correct labels for all images in I. This is a strong assumption, as a substantial fraction of images are hard to label for humans if not primed. The rationale is that a motivated attacker may spend substantial effort to label the images, automated image search facilities might reveal the source image, or algorithmic classifiers may be able to label images. (We are unaware of any algorithm that can identify objects in Mooney images, but we cannot guarantee that such algorithm does not exist; thus we assume the attacker (artificially) knows all labels.)
2) We assume the adversary knows the probabilities n_i and p_i. While knowing the exact values requires substantial work by the attacker (basically replicating our study), getting approximations is relatively easy, and one should not rely on an assumed bound on their correctness.
3) The adversary is free to answer the questions at any time, i. e., the answer times can be freely manipulated. (Even though the adversary cannot gain any advantage from this with the current prototype, this may be relevant for alternative implementations that more carefully take the answer time into account.)

Consequently, the security of the scheme solely relies on the partition of the shown images into the primed and non-primed images, i. e., the sets I_P and I_N.

E. Static Scoring

One straightforward scoring strategy, used by Denning et al. [16], is what we call *static scoring*. We briefly describe static scoring here so later we can compare our new scoring strategy, *dynamic scoring*, with it. There are four basic events that can occur for a single image:

- A primed image (with index i) is
  - labeled correctly:
    Occurs with probability p_i, assigned score $s_{p,c}$.
  - labeled incorrectly:
    Occurs with probability $1 - p_i$, assigned score $s_{p,f}$.
- A non-primed image (with index i) is
  - labeled correctly:
    Occurs with probability n_i, assigned score $s_{n,c}$.
  - labeled incorrectly:
    Occurs with probability $1 - n_i$, assigned score $s_{n,f}$.

Now static scoring assigns the value 1 to the two "good" events, i. e., $s_{p,c} = 1$, $s_{n,f} = 1$ and 0 to the two "bad" events $s_{p,f} = 0$, $s_{n,c} = 0$. In other words, this scoring strategy counts the "good" events that happened.

F. Dynamic Scoring

Static scoring does not differentiate between different probability values, thus loses information. We propose an alternative method, *dynamic scoring*, which takes inspiration from the notion of *self-information* or *surprisal*, a well-known concept in information theory [40]. Self-information denotes the information content associated with a single event, as opposed to entropy which is a property of an entire distribution.

The self-information $I(E^*)$ of an event E^* with probability p_i is defined as

$$I(E^*) = -\log(p_i),$$

where we use logarithms to base e throughout this work. For dynamic scoring, score each event with its surprisal, i. e.,

$$s_{p,c} = \ln(p_i), \qquad s_{p,f} = \ln(1 - p_i),$$
$$s_{n,c} = \ln(n_i), \qquad s_{n,f} = \ln(1 - n_i).$$

Note that we invert the sign of $I(E^*)$ so that a higher score refers to a better match, i. e., "less surprisal". Consequently, the scores are negative.

For an intuition on why dynamic scoring improves on static scoring consider the event E^* that the user wrongly labels a primed image. Let us assume a fixed "priming effect", i. e., the difference $d_i = p_i - n_i = 0.5$ is constant. We first consider an image where $p_i = 0.5$ (and thus $n_i = 0$), i. e., the

primed image is labeled correctly and incorrectly with the same probability. Then the event E^* does carry little information, as it is a plausible outcome for a legitimate (primed) user. Second, we consider the case where $p_i = 1$ (and thus $n_i = 0.5$), then every primed image will be labeled correctly by the legitimate (primed) user. Thus, if event E^* happens, we can be certain that it's not the legitimate user participating in the protocol. Static scoring gives the same score 0 in both cases, while dynamic scoring gives a score of $-\infty$ and thus indicates that this event can only be caused by an impostor.

Legitimate (Primed) User Score. For the legitimate user, the expected value of the score S_i for a single image with index i equals

$$\begin{aligned} \mathrm{E}(S_i) &= \frac{1}{2} \cdot (p_i \cdot \ln(p_i) + (1 - p_i) \cdot \ln(1 - p_i)) \\ &\quad + \frac{1}{2} \cdot (n_i \cdot \ln(n_i) + (1 - n_i) \cdot \ln(1 - n_i)) \,, \end{aligned}$$

which equals the average of the Shannon entropies of Bernoulli-distributed random variables B_{1,p_i} and B_{1,n_i} with mean p_i and n_i, respectively,

$$\mathrm{E}(S_i) = \frac{1}{2} \left(\mathrm{H}(B_{1,p_i}) + \mathrm{H}(B_{1,n_i}) \right) ,$$

where $H(X)$ denotes the Shannon entropy, which is the expected value of the self-information $H(X) = E[I(X)]$.

Adversary Score. The adversary does not know whether the image was primed or not (this is what the security of the scheme rests on). Recall that we assume the adversary knows the labels and knows the probabilities p_i and n_i. We assume that the same number of images is primed and non-primed so that a single random image is primed with probability $\frac{1}{2}$. An adversary can basically decide to give the correct label or the wrong label, based on the known probabilities. If the adversary gives the correct label, the score (for that single image) will be

$$(s_{p,c} + s_{n,c})/2 = (\ln(p_i) + \ln(n_i))/2,$$

and if the adversary gives the incorrect label the score will be

$$(s_{p,f} + s_{n,f})/2 = (\ln(1 - p_i) + \ln(1 - n_i))/2.$$

So an adversary can calculate both values and pick the one that has a higher expected score.

IV. Experiment 1: Pre-Study

A pre-study was performed to identify the critical parameters of the MooneyAuth scheme. The critical parameters are: (1) p_i, the probability that a primed image is correctly labeled and (2) n_i, the probability that a non-primed image is labeled correctly. From these values, we can then derive (3) the size k of the sets I_P and I_N. These parameters were then used in the following experiments.

While there is no ethics committee covering this type of studies at the organizations involved in this research, there are strict laws and privacy regulations in place that must be obeyed. The experiments comply with these strict regulations. The data we collected about a participant cannot be linked back to a respondent, as the data is in quite broad categories only. We did not collect any personal identifiers (IP address, device identifier, name, or similar), and did not use third-party components that may still log such data. Before any data was recorded, the respondents were informed about the purpose of the experiment and how the contributed data will be managed, and that they can leave the experiment at any time.

A. Experimental Setup

We used a total of 120 images. For each participant, we used 10 primed images $|I_P| = 10$ and 20 non-primed images $|I_N| = 20$, i. e., an asymmetric distribution of primed and non-primed images, randomly selected from the 120 images. Choosing $|I_N|$ to be larger than $|I_P|$ helps to speedup the enrollment process. We developed a web application to conduct the experiment and measured the parameters p_i, n_i.

Enrollment (Priming) Phase. For the enrollment phase, a random subset of $|I_P| = 10$ Mooney images was selected for each participant. Priming consisted of four steps:

(i) *Introduction*: The experiment started with a brief introduction and explanation of how the experiment will proceed. We provided participants with the necessary written explanation on the web page that this study was about an alternative web-based authentication scheme. Participants were informed about the two experimental phases (enrollment and authentication). They were further informed to be contacted via email, after the enrollment phase, to take part in the authentication phase. Besides, we provided a link with further information about Mooney images and implicit memory for the interested participant.

(ii) *Priming* 1: For each image from the subset I_P, we first presented the Mooney image for 3.5 seconds, then the original gray-scale image for another 3.5 seconds, then again the Mooney image. To make the shifting between the images more comprehensible, we gradually transitioned between the images, i. e., fading out the first image, while fading in the second image. A label (a single English word) that described the hidden object in the image was displayed during the original gray-scale image presentation. We consider this approach a reasonable tradeoff between giving enough time to prime the image and spending time on the enrollment process.

(iii) *Survey*: After the first priming phase, the participants were asked to fill out a short questionnaire with basic questions such as age, field of work, gender, and opinion about the usability of current web authentication systems. This survey was intended to provide the participants with a short break before the second priming phase. In addition, we used the data collected from this survey for a statistical assessment of the participants.

(iv) *Priming* 2: In the second priming phase, we repeated the first priming phase for the same 10 images in a new pseudo-randomized order. Overall, users saw each Mooney image and its corresponding gray-scale image twice.

Authentication (Recall) Phase. Participants were invited via email to take part in the authentication phase. Each participant was provided with an individual link. In order to test how

long the effects of priming and authentication performance lasted, we performed the authentication phase in two separate groups at two different points in time (approximately two weeks apart). The authentication phase was composed of two main steps:

(i) *Introduction*: Before the authentication started the task was described. Each participant was asked to view the Mooney image, and to label the hidden object in the image as fast as possible. Participants were specifically asked to label each image using a single English word. Importantly, participants were asked to label the images regardless of what they have seen in the priming phase. If the participants could not identify the hidden object (possibly because this image was not used in the priming phase), they were asked to press the "I don't know" button. These instructions were provided in a written form on the web page.

(ii) *Authentication*: For each returning participant, we selected a subset $I_N \subset I \setminus I_P$ of size $|I_N| = 20$. All Mooney images from the entire set $I_P \cup I_N$ were presented to the participants for labeling in random order. The interface used for this labeling task can be seen in Figure 2.

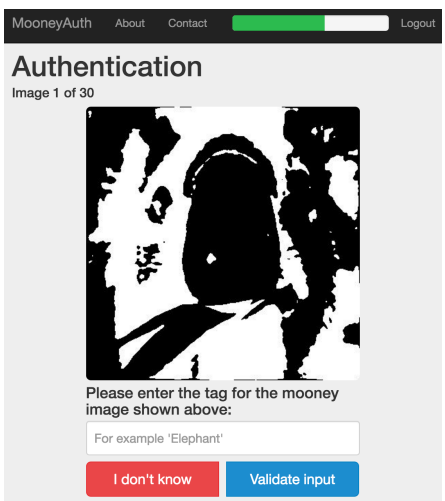

Fig. 2. Screenshot of the user interface during the authentication phase.

Please note that the website as used in Experiment 1 had a bug, which led to a layout change caused by an information banner fading out during the labeling process in the authentication phase. This could have led participants to click the "I don't know" button accidentally instead of selecting the text entry field. It seems very unlikely that this bug has affected the results: we have not received any feedback from the participants mentioning this issue, the fading out related miss-click could have only occurred in specific instances with a slow Internet connections, and when we filtered the participants that may have been affected based on the text input time the overall results even slightly improved. Furthermore, we fixed this potential issue for Experiments 2 and 3, and these report very similar results. This confirms that the bug had minimal or no influence on the results.

B. Implementation

To perform the experiments, we implemented a web application based on the *Model, View, Controller* (MVC) design pattern. The front-end (View) is based on the Bootstrap framework to accelerate development, the back-end (Model and Controller) is written in PHP, and data is stored in a MySQL database. To compute an edit distance during the authentication phase, we used a C implementation of the Damerau-Levenshtein algorithm which was included as external PHP module. Data was transmitted using transport layer security (TLS) to protect the privacy of the participants. To be compliant with the federal data protection act and privacy laws, users were informed about what data was collected and had to consent to the processing and storing of the data. Collected data was stored in encrypted form. We used the free web analytics software Piwik on the web server to derive statistics about the web application's usage. Every user was able to opt-out and the usage of the *Do Not Track* (DNT) HTTP header was honored.

C. Matching Labels

For each image, we created a small set of correct labels (typically two to five labels). All labels were converted to lowercase before comparison. We computed the Damerau-Levenshtein distance (string edit distance considering insertion, deletion, substitution, and transposition of adjacent letters) between the provided label and all given labels for that image. If one label had a distance less or equal to 1, we marked it to be correct. This ensures that a variety of typical deviations is accepted, such as simple spelling errors, plural endings, British/American spelling differences, and such.

Although, the use of an open text field to provide answers has drawbacks considering entry time and error rate (especially on mobile devices), we decided not to use alternative methods such as selecting the correct answer from multiple choice. Previous work has shown that using multiple choice answers leads to higher recognition rates for non-primed Mooney images [28]. First, the number of choices gives us a lower bound for the n_i and second providing a choice of labels already exhibits priming effects.

D. User Participation

Participants were recruited via several email distribution lists and social media. To motivate participants, we raffled gift cards to those who finished both phases. For this experiment, 360 people started the enrollment phase. We sent out 323 invite emails for the authentication phase because 37 participants had not finished their enrollment (6 stopped at the introduction tutorial, 16 during the first priming, 6 during the survey, 9 during the second priming). From those re-invited to the authentication phase 230 finished, 6 started but have not finished, and 87 never tried to start the phase. A high dropout rate between enrollment and authentication was expected, as we have not verified email addresses during the enrollment of Experiment 1 nor have we filtered obviously fake email addresses. Furthermore, misclassification of our invite email as spam might have occurred, as well, which would explain the high number of users that not even tried to start the authentication phase.

We collected, with users' consent, basic statistics such as country of origin and timing from the server logs, as well as the results from a survey; a summary of the statistics can be found in Appendix A. Please note, the reported numbers of the questionnaire in the appendix differ from the actual number of participants, as providing answers was not mandatory. About four out of five participants were between 20 and 30 years old, but all age groups were represented, and about four out of five were male. Most participants were from France and Germany due to the mailing lists we used, but people from over 30 countries participated. The majority of them liked to use the MooneyAuth scheme.

As a result of the sampling process, the participants in this and the following experiments are skewed towards young and male participants working in the sciences. Previous work found no evidence indicating differences in recognition rates of Mooney images for gender or occupation of the primed participants [25], [26], [27], [28]. Unfortunately, there is no data available on priming effects in different age groups.

E. Results

We now present the results of our first experiment that helped us to estimate and test parameters (e. g., labeling).

1) Estimating p_i, n_i, and d_i: The main result of Experiment 1 is the estimation of the parameters p_i and n_i for the tested images. We find that the average difference d over the individual $d_i = p_i - n_i$, which is a good indicator for the overall performance, is 0.43. This is a fundamental improvement over the previous work [16], which achieved an average difference of $d = 0.07$.

A more detailed view is given in the plot in Figure 3, which shows these parameters for each individual image. Each data point indicates one image, with the positions on the x-axis (y-axis) representing the empirical values for p_i (n_i). The plot shows our main result for the full dataset. To improve comparability with previous findings, we printed our results as an overlay on top of the plot from previous work [16]. One might reason that the compared time frames are not the same (20 days and 28 days). However, we show in Section V that our Mooney image priming effect declines only moderately over time allowing one to consider this a fair comparison.

The (diagonal) lines are intended to help in the comparison of the results in the layered graph. The small solid line in the top of the graph ($p_i = 0.07 + n_i$) indicates the average value of the difference d of the previous work [16], it corresponds to the bold solid line in the bottom of the graph. This line represents the average for our system ($p_i = 0.43 + n_i$), while the third solid line ($p_i = 0.5 + n_i$) indicates the line with $d_i = 0.5$.

2) Response Time: A summary is given in Table II. The average time to label an image is around 10 seconds with a high standard deviation. (Maximum timing can be more than 10 minutes). Median values are more robust to outliers. They are closely grouped together ($7.25 - 7.89$ seconds). The only exception can be seen in the correctly labeled primed images. These images were substantially faster (a median of 6.30 seconds).

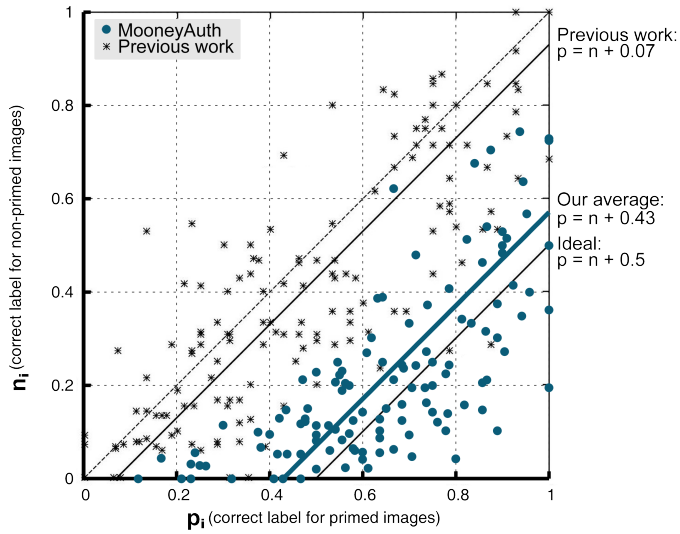

Fig. 3. Priming effect comparison: p_i versus n_i plot for our scheme after 20 days (points, blue) and previous work [16] after 28 days (stars, black).

3) Strict vs. Relaxed Labeling: The way we use for testing the labels for correctness may obviously affect the measured values (and thus the performance of the scheme). To evaluate if the strict labeling, as described in Section IV-C, gives reasonable results, or whether more sophisticated measures (e. g., a lexical database that includes *synsets* to find related words) needed to be taken, we additionally assessed the quality of the comparison by hand. We tested all labels that were classified as "wrong" in the automatic test. In this manual "clean up session", we added some labels to the set of accepted labels that were synonymous to existing labels, which we missed in the original creation of the labels (e. g., we added "carafe" for an image showing a "pitcher"), we added some generalized terms (e. g., "animal" instead of "tiger"), and very similar species that were easy to confuse in the images (e. g., "bee" and "ant"). We grouped those labels as "similar", and everything else as "wrong" as before.

Contrary to our expectation, relaxed labeling slightly worsens the performance. While for strict labeling we have $d = 0.43$, for the relaxed labeling we have $d = 0.42$, a small but noticeable difference. This might be explained by the fact that some "similar" cases, in particular generalizations, are so general that they can be guessed (e. g., 77 of the 120 images were showing animals). Consequently, in all following studies we used the strict labeling, which in addition is computable without human intervention.

V. Experiment 2: Long-term Behavior Study

It is well-known that, in principle, priming can last over very long times [13]. However, this is not known for priming on Mooney images. In a second experiment, we measured the long-term effects of the priming.

A. Experimental Setup

This experiment is an extension of the first experiment. It extends Experiment 1 in two aspects: first, we divided the data gathered in Experiment 1 into two batches by the

TABLE I. STATISTICS ON THE DURATION AND AVERAGE EVENT PROBABILITY PER EXPERIMENT.

	Duration (in days)			Results		
	Mean	SD	Median	Average p_i	Average n_i	Average d_i
Experiment 1	18.0	8.8	20	0.648	0.219	0.429
Experiment 2						
– Batch 1	8.7	2.2	9	0.726	0.226	0.500
– Batch 2	25.1	4.2	25	0.586	0.215	0.371
– Batch 3	264.3	3.8	264	0.499	0.252	0.247
Experiment 3	19.9	4.7	21	0.642	0.203	0.439

TABLE II. STATISTICS ON THE TIMING FOR THE LABELING OPERATION, ALL VALUES ARE IN SECONDS.

Class	Median	Mean	SD
Primed/correct	6.30	8.62	8.76
Primed/false	7.25	10.24	12.54
Non-primed/correct	7.89	11.17	24.83
Non-primed/false	7.30	9.55	15.38

time between enrollment and authentication; second, we re-invited the participants of Experiment 1 after approximately 8.5 months again and measured the p_i, n_i decline over time. Therefore, we can compare three different batches (9, 25, and 264 days), details are listed in Table I.

B. User Participation

People from the first batch were invited to authentication approximately 10 days after the first invitation to the enrollment, people from the second batch after about three and a half weeks. For each participant, we measured the time between priming and authentication. For the first batch, this difference has a median of 9 days, for the second batch, it has a median of 25 days. For the third batch, the median is 264 days. Further details on the participants are given in the Appendix A.

C. Results

Detailed information is provided in Figure 4 and Table I. We see a moderate decline of the priming effect over the first couple of weeks: the average value of the d_i is 0.500 for the first batch and 0.371 for the second batch, both for strict labeling. However, over longer times, the decline becomes much less pronounced; in fact, 264 days after the initial priming we still measure an average d_i of 0.247.

This is shown in more detail in Figure 4, which shows scatter plots for p_i and n_i, separated for the first batch (top), second batch (center), and third batch (bottom). We can additionally see that even in the third batch, there is a substantial number of images with a d_i greater than 0.5 (dashed line on the lower right).

Additionally, Table I shows the average p_i and n_i for each batch. As expected, the values for n_i do not vary over time (as no priming took place), but the values for p_i do change.

VI. EXPERIMENT 3: MOONEYAUTH STUDY

Based on the findings of our first study, we conducted a third study with the estimated parameters p_i and n_i. This experiment is designed as a realistic test of the overall performance of the authentication scheme.

A. Experimental Setup

The experimental setup was very similar to the setup for our first experiment as described in Section IV-A.

The main difference is the reduced set of images. We used a subset of 20 images of the original image database, the same subset for all users, and computed a random partition of this reduced database for each participant. We selected those images with the best performance in the first experiment, i. e., those images with the highest values $d_i = p_i - n_i$. The selected images had values d_i between 0.79 and 0.57, on average 0.643. For each user, we used 10 primed and 10 non-primed images, i. e., $|I_P| = |I_N| = 10$.

There were no changes to the enrollment phase. The authentication phase worked as before, but as we learned in Experiment 1 that the strict labeling outperforms the relaxed labeling, we only used the strict labeling. The goal of this experiment was to evaluate the suitability of the authentication method, including potential cross-contamination of the memory when several images with good priming effects were learned by a single user (an effect we could not study in the first experiment). Also, the measured and presented statistics are tailored towards this goal.

B. User Participation

Participants were recruited via email distribution lists. We took several measures to avoid that participants of the first or second experiment also participated in the third: we used (mostly) disjoint mailing lists, asked users in the questionnaire if they participated before, filtered all emails used for the login that have participated in the first, and placed a cookie that allowed us to detect multiple participations. Participating in both studies has to be prevented as the images in the third experiment are a subset from the first experiment, so being primed on some images in the first experiment can disturb the results of the third experiment. However, the effect of duplicate participants is small, as the overlap of primed images and the 20 images in the third experiment is less than two on average. Again, we raffled gift cards to those who finished both phases.

About half of the 70 participants in this experiment were between 20 and 30 years, but all age groups were represented. About 3 out of 4 were male. Most participants were from France and Germany, because of the mailing lists we used. The results of the questionnaire are shown in Appendix B.

C. Results

The main result of this experiment is a precise estimation of the performance of the proposed authentication scheme.

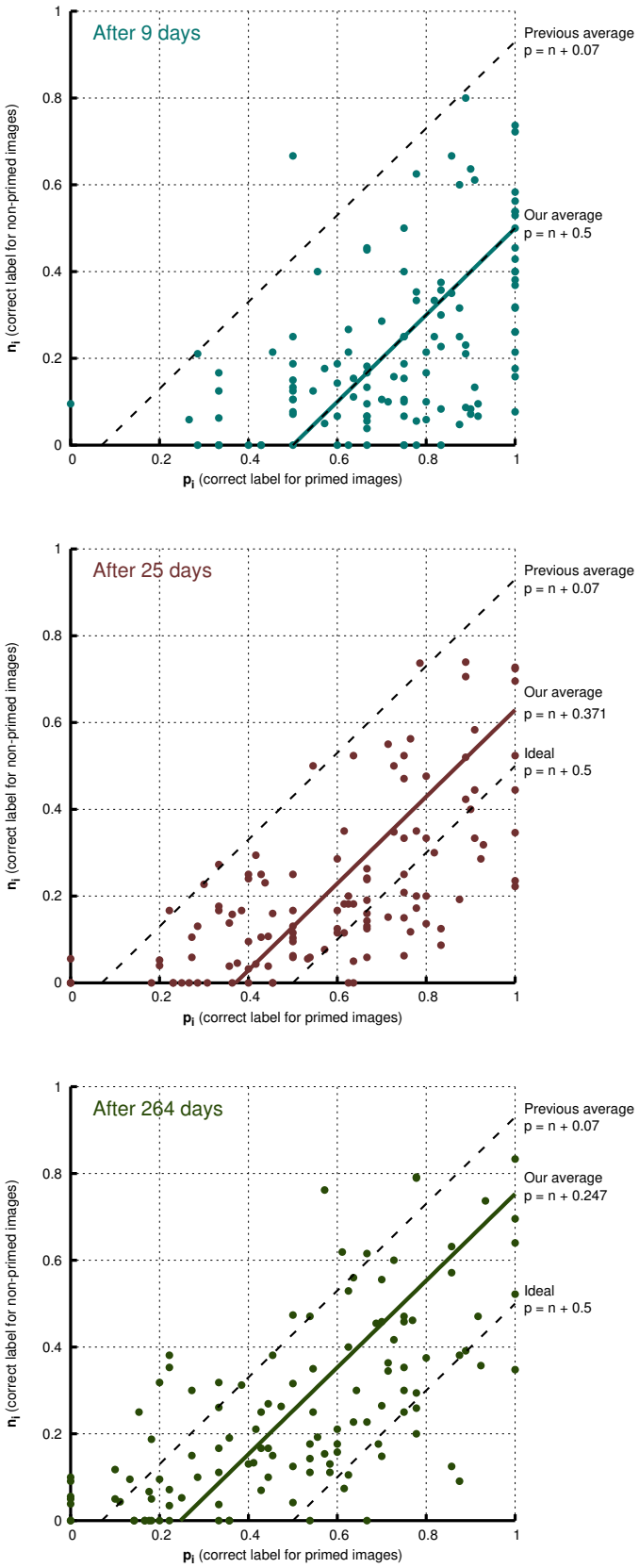

Fig. 4. Priming effect decline over time: p_i versus n_i plot for the first batch after 9 days (top), the second batch after 25 days (center), and the third batch after 264 days (bottom).

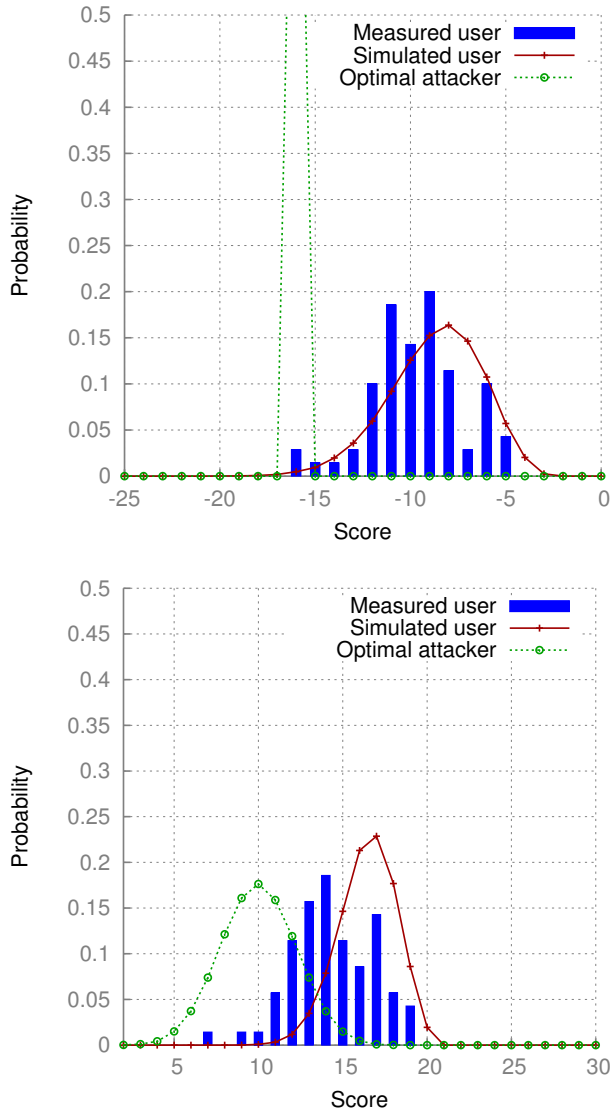

Fig. 5. Distribution of measured (bar, blue) and estimated (solid line, red) scores for dynamic scoring (top) and static scoring (bottom).

In addition, we compare the static and the dynamic scoring strategy.

1) Performance: The complete graphs illustrating the distribution of scores are shown in Figure 5, both for dynamic scoring (top), and static scoring (bottom). The x-axes give the scores assigned to a run (rounded to integers if necessary), and the y-axes the relative frequency. The blue bars give the actual measured distribution determined in the third experiment, while the red solid line gives the estimated distribution of score values for a legitimate user (see Section VI-C2, using the estimated parameters p_i and n_i from above). The green dashed line gives the distribution of an impostor using the optimal strategy as described in Section III-F.

We measure the performance of the scheme in terms of the false acceptance and false reject rates. The false acceptance rate (FAR) is an indicator for the security of the protocol; it gives the likelihood that an impostor is (falsely) classified as a legitimate user, i. e., "accepted". For fallback authentication

TABLE III. PERFORMANCE OF THE SCHEME FOR PARAMETERS $|I_P| = |I_N| = 10$.

	Target FAR	Score Thres.	Resulting FRR Sim.	Resulting FRR Meas.
Static scoring	0.1 %	17	48.9 %	76 %
	0.5 %	16	27.6 %	67 %
	1.0 %	15	13.0 %	56 %
Dynamic scoring	0.1 %	-16	0.30 %	2.86 %
	0.5 %	-16	0.30 %	2.86 %
	1.0 %	-16	0.30 %	2.86 %

TABLE IV. STATISTICS ON THE OVERALL TIMING (IN SECONDS) FOR EXPERIMENT 3.

	Mean	Median	Max	Min	SD	Var
Enroll - Tutorial	24	23	58	13	8	58
Enroll - Priming 1	113	107	170	99	15	221
Enroll - Survey	51	47	143	23	20	390
Enroll - Priming 2	105	101	186	94	14	196
Total - Enroll	294	**(5 minutes)**				
Auth - Tutorial	28	24	97	3	16	262
Auth - Labeling	177	163	472	70	75	5597
Total - Auth	206	**(3.5 minutes)**				

schemes (which can apply strict rate-limiting and other techniques to limit the capabilities of an impostor) FARs in the range of 0.01 and 0.001 can be considered acceptable (Denning et al. [16] considered a FAR of 0.005). For a given FAR, we can determine the threshold that meets this FAR, which provides us with the false reject rate (FRR), i. e., the probability that a legitimate user is denied access to the system. Denning et al. [16] considered an FRR of 0.025 to be acceptable.

Figure 5 and Table III depict the basic performance of the proposed scheme. We can see that for the dynamic scoring, the scheme achieves simulated FRRs of $0.3\,\%$ for FARs between $1\,\%$ and $0.1\,\%$, and measured FRRs of $2.86\,\%$. (While it may be surprising that the measured FRR are higher than the simulated FRRs, please note that only two participants achieved a dynamic score of -16, which are solely responsible for the relatively high FRR.) Still, an FRR of $2.86\,\%$ is pretty much within the bounds of previous work.

Some statistics about the duration of the experiments and the properties of the used Mooney images are summarized in Table I. Some statistics about the duration of each phase is given in Table IV. For example, it shows that the enrollment phase took 5.0 min on average (including tutorial and questionnaire), and the authentication phase 3.5 min.

2) The Simulation: Besides the measured data from the user experiment, we use simulated numbers to provide additional insights. These simulations are based on the estimated parameters p_i, n_i determined in the first experiment, where we selected the 20 best images and used those p_i, n_i. We simulated $100\,000$ authentication attempts as follows:

- Choose random subsets I_P and I_N from the available images.
- Simulate a user (primed on I_P) logging in, based on the collected probabilities p_i, n_i, and compute the score.
- Simulate an optimal adversary (as defined above), and compute the score.

An interesting observation is that the simulation based on the probability values from the previous experiment is relatively accurate. We can see that the shape of the simulated distribution (red solid line) closely resembles the shape of the measured distribution (blue bars). The only substantial difference is that the distribution is shifted towards lower values, i. e., the mean changes from -8.45 to -9.6 (for the dynamic scoring), and from 16.5 to 14.4 (for the static scoring). In other words, the performance we measured is slightly worse than predicted by the simulation, which can have several plausible reasons: (i) The time difference between enrollment and authentication for Experiment 1 (when estimating the d_i) was slightly shorter than for Experiment 3 (mean duration of approx. 18 days vs. 20 days). (ii) Being primed on several images with good priming properties in parallel may cross-contaminate the participant's memory and thus worsen the overall recall. However, from this experiment we see that even if this effect plays a role, its influence is relatively small.

We can also see that the dynamic scoring substantially outperforms the static scoring. Table III lists, for several target FARs, the resulting FRRs, both for dynamic and static scoring. We see that for all listed FARs, the resulting FRRs are substantially better with dynamic scoring, both for the measured data and for the simulated values.

VII. SECURITY ANALYSIS / DISCUSSION

In the proposed authentication scheme, the priming effect of Mooney images is used to help users memorizing their authentication secret, using implicit instead of explicit memory. However, it is important to note that, similarly to graphical authentication schemes based on explicit memory, the security of this scheme relies on the subset I_P only, and does not depend on the properties of Mooney images. In fact, our security model considers a powerful attacker who (artificially) knows the solution (label) for every image and still fails to authenticate. (There is an indirect dependency, however, as a weak priming effect will typically be compensated by a lower threshold to control the false acceptance rate and thus make attacks easier.) Once a catalog of images with good priming properties is used (e. g., $d_i > 0.5$, see Figure 6), the scheme is resilient to rate limited guessing attacks. Note that all users of the authenticating service share the same set I of such images ($I_P, I_N \subset I$). Thus, selecting the images is a one-time task.

The secret used for authentication is the set of primed images I_P, which is a subset of all images presented to the user in the authentication phase $I_P \cup I_N$. Effectively, I_P is a randomly chosen subset, so there is no bias of user choice involved (in contrast to passwords and many other schemes), which facilitates the security analysis. The authentication score computed by our scheme is not only based on the primed images the user can identify, but also on the non-primed images that an adversary is not able to determine. As a consequence, an adversary that can decode Mooney images without going through the priming phase has no advantage for breaking the security of the proposed scheme, if it is unknown on which images the victim was primed on. Also, a user connecting to the server under a false username and obtaining the presented images does not affect the security. To avoid intersection attacks, it is mandatory that the same set of non-primed images I_N is presented at each login attempt.

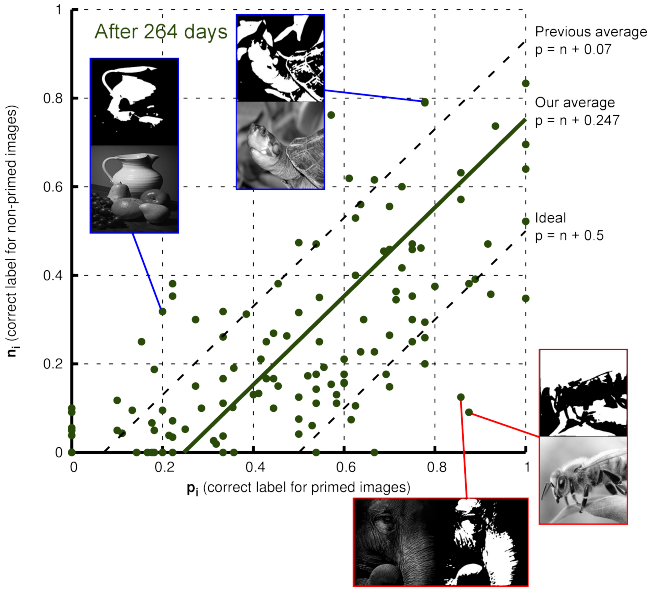

Fig. 6. Example images with longtime (264 days) low (top, blue) and high (bottom, red) priming effects.

Just like most other schemes, our scheme is susceptible to phishing attacks: An attacker can query the authentication server for the images, present them to the legitimate user, record the timings, and replay those to the server. All standard measures to prevent phishing attacks apply here as well. Furthermore, an *active phishing attack* is required, i. e., the attacker needs to query the server to get the correct set of images, which may be detected on the server's side.

While passwords can be stored (relatively) secure on the login server using iterated password hashes and random salts to prevent guessing attacks, this is not feasible for a large range of fallback authentication methods, e. g., as for knowledge questions also approximate answers should be counted, which typically requires storing the solution in plaintext. Similar, there is no (obvious) way to store the secret information (i. e., the indices describing the set I_P for our scheme).

Guessing attacks against our scheme can be avoided just as for other fallback authentication schemes. For example, by putting substantial limits on the guessing rate (e. g., one attempt per day), and a lock-out period, i. e., if account recovery is initiated, the original owner is notified, e. g., via the stored mail, and has 24 hours to abort the recovery if it was started by somebody else. All of these measures are implemented for other schemes as well.

We did not study interference properties, i. e., how well a user can remember a secret if the user is using the system on several servers in parallel (with different sets of primed images I_P). However, most smaller websites use fallback authentication by email or use a single-sign-on solution. Thus a more involved fallback authentication scheme like ours will mostly be of interest to large email providers or social network sites, and thus a user will only use very few parallel instances.

Our experiments are conducted on a limited set of 20 Mooney images out of which 10 were primed. This might open the question whether retrieval of primed Mooney images can get harder when more images are primed. An experiment by Ludmer et al. has used Mooney images to explore memory retrieval in the human brain by priming users with 30 randomly selected images [28]. They have shown that if the solution to a primed Mooney image is retained one week after priming, it is essentially retained to the same degree also three weeks afterward. This suggests that even when a larger sample size of Mooney images is used retrieval of primed Mooney images are likely to retain after longer periods of time.

VIII. Conclusion

Authentication schemes based on implicit memory relieve the user of the burden of actively remembering a secret (such as a complicated password). This paper presents a new implicit memory-based authentication scheme that significantly improves previous work by using a more efficient imprinting mechanism, namely Mooney images, and optimizing the scoring mechanism. We implemented a comprehensive prototype and analyzed the performance and security of our proposal in a series of experiments. Results are promising and show that our scheme is particular suited for applications where timing is not overly critical, such as fallback authentication.

Fig. 7. Mooney image and the corresponding original image.

Acknowledgment

Maximilian Golla is supported by the German Research Foundation (DFG) Research Training Group GRK 1817/1.

References

[1] A. Adams and M. A. Sasse, "Users Are Not the Enemy," *Communications of the ACM*, vol. 42, no. 12, pp. 40–46, Dec. 1999.

[2] M. D. Barense, J. K. W. Ngo, L. H. T. Hung, and M. A. Peterson, "Interactions of Memory and Perception in Amnesia: The Figure-Ground Perspective," *Cerebral Cortex*, vol. 22, no. 11, pp. 2680–2691, Nov. 2012.

[3] C. J. Berry, D. R. Shanks, M. Speekenbrink, and R. N. A. Henson, "Models of Recognition, Repetition Priming, and Fluency: Exploring a New Framework," *Psychological Review*, vol. 119, no. 1, pp. 40–79, Jan. 2012.

[4] R. Biddle, S. Chiasson, and P. Van Oorschot, "Graphical Passwords: Learning from the First Twelve Years," *ACM Computing Surveys*, vol. 44, no. 4, pp. 19:1–19:41, Aug. 2012.

[5] J. Blocki, S. Komanduri, L. F. Cranor, and A. Datta, "Spaced Repetition and Mnemonics Enable Recall of Multiple Strong Passwords," in *Symposium on Network and Distributed System Security*, ser. NDSS '15. San Diego, California, USA: The Internet Society, Feb. 2015.

[6] H. Bojinov, D. Sanchez, P. Reber, D. Boneh, and P. Lincoln, "Neuroscience Meets Cryptography: Designing Crypto Primitives Secure Against Rubber Hose Attacks," in *USENIX Security Symposium*, ser. SSYM '12. Bellevue, Washington, USA: USENIX Association, Aug. 2012, pp. 129–141.

[7] J. Bonneau, E. Bursztein, I. Caron, R. Jackson, and M. Williamson, "Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google," in *International World Wide Web Conference*, ser. WWW '15. Florence, Italy: ACM, May 2015, pp. 141–150.

[8] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes," in *IEEE Security and Privacy*, ser. SP '12. San Jose, CA, USA: IEEE, May 2012, pp. 553–567.

[9] J. Bonneau, M. Just, and G. Matthews, "What's in a Name? Evaluating Statistical Attacks on Personal Knowledge Questions," in *Financial Cryptography and Data Security*, ser. FC '10. Tenerife, Canary Islands, Spain: Springer, Jan. 2010, pp. 98–113.

[10] J. Bonneau and S. Preibusch, "The Password Thicket: Technical and Market Failures in Human Authentication on the Web," in *Workshop on the Economics of Information Security*, ser. WEIS '10, Cambridge, Massachusetts, USA, Jun. 2010.

[11] J. Bonneau and S. Schechter, "Towards Reliable Storage of 56-bit Secrets in Human Memory," in *USENIX Security Symposium*, ser. SSYM '14. San Diego, California, USA: USENIX Association, Aug. 2014, pp. 607–623.

[12] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung, "Fourth-Factor Authentication: Somebody You Know," in *ACM Conference on Computer and Communications Security*, ser. CCS '06. Alexandria, Virginia, USA: ACM, Nov. 2006, pp. 168–178.

[13] C. B. Cave, "Very Long-Lasting Priming in Picture Naming," *Psychological Science*, vol. 8, no. 4, pp. 322–325, Jul. 1997.

[14] D. Davis, F. Monrose, and M. K. Reiter, "On User Choice in Graphical Password Schemes," in *USENIX Security Symposium*, ser. SSYM '04. San Diego, California, USA: USENIX Association, Aug. 2004, pp. 151–164.

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '09. Miami, Florida, USA: IEEE, Jun. 2009, pp. 248–255.

[16] T. Denning, K. Bowers, M. van Dijk, and A. Juels, "Exploring Implicit Memory for Painless Password Recovery," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. Vancouver, British Columbia, Canada: ACM, May 2011, pp. 2615–2618.

[17] R. J. Dolan, G. R. Fink, E. Rolls, M. Booth, A. Holmes, R. S. Frackowiak, and K. J. Friston, "How the Brain Learns to See Objects and Faces in an Impoverished Context," *Nature*, vol. 389, no. 6651, pp. 596–599, Oct. 1997.

[18] S. L. Garfinkel, "Email-Based Identification and Authentication: An Alternative to PKI?" *IEEE Security and Privacy*, vol. 1, no. 6, pp. 20–26, Dec. 2003.

[19] M. S. Gazzaniga, *Cognitive Neuroscience: The Biology of the Mind*, 4th ed. New York, NY, USA: W. W. Norton & Company, Inc, 2013.

[20] M. A. Gluck, *Learning and Memory: From Brain to Behavior*, 2nd ed. New York, NY, USA: Worth Publishers, 2014.

[21] V. Griffith and M. Jakobsson, "Messin' with Texas: Deriving Mother's Maiden Names Using Public Records," in *Conference on Applied Cryptography and Network Security*, ser. ACNS '05. New York, NY, USA: Springer, Mar. 2005, pp. 91–103.

[22] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig, "Use Your Illusion: Secure Authentication Usable Anywhere," in *USENIX Symposium on Usable Privacy and Security*, ser. SOUPS '08. Pittsburgh, PA, USA: ACM, Jul. 2008, pp. 35–45.

[23] A. Hern, "Google Aims to Kill Passwords," May 2016, https://www.theguardian.com/technology/2016/may/24/google-passwords-android, as of December 20, 2016.

[24] P.-J. Hsieh, E. Vul, and N. Kanwisher, "Recognition Alters the Spatial Pattern of fMRI Activation in Early Retinotopic Cortex," *Journal of Neurophysiology*, vol. 103, no. 3, pp. 1501–1507, Jan. 2010.

[25] F. Imamoglu, T. Kahnt, C. Koch, and J.-D. Haynes, "Changes in Functional Connectivity Support Conscious Object Recognition," *NeuroImage*, vol. 63, no. 4, pp. 1909–1917, Dec. 2012.

[26] F. Imamoglu, C. Koch, and J.-D. Haynes, "MoonBase: Generating a Database of Two-Tone Mooney Images," *Journal of Vision*, vol. 13, no. 9, pp. 50–50, Jul. 2013.

[27] J. M. Kizilirmak, J. Galvao Gomes da Silva, F. Imamoglu, and A. Richardson-Klavehn, "Generation and the subjective feeling of "aha!" are independently related to learning from insight," *Psychological Research*, vol. 80, no. 6, pp. 1059–1074, Aug. 2016.

[28] R. Ludmer, Y. Dudai, and N. Rubin, "Uncovering Camouflage: Amygdala Activation Predicts Long-Term Memory of Induced Perceptual Insight," *Neuron*, vol. 69, no. 5, pp. 1002–1014, Mar. 2011.

[29] D. E. Meyer and R. W. Schvaneveldt, "Facilitation in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations," *Journal of Experimental Psychology*, vol. 90, no. 2, pp. 227–234, Oct. 1971.

[30] D. B. Mitchell, A. S. Brown, and D. R. Murphy, "Dissociations Between Procedural and Episodic Memory: Effects of Time and Aging," *Psychology and Aging*, vol. 5, no. 2, pp. 264–276, Jun. 1990.

[31] C. M. Mooney, "Age in the Development of Closure Ability in Children," *Canadian Journal of Psychology*, vol. 11, no. 4, pp. 219–226, Dec. 1957.

[32] R. Nieva, "Yahoo Wants to Kill Passwords," Oct. 2015, http://www.cnet.com/news/yahoo-wants-to-kill-passwords-with-revamped-mail-app/, as of December 20, 2016.

[33] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[34] Real User Corporation, "The Science Behind Passfaces," Jun. 2004, http://www.realuser.com/published/ScienceBehindPassfaces.pdf, as of December 20, 2016.

[35] K. Renaud and M. Just, "Pictures or Questions? Examining User Responses to Association-Based Authentication," in *BCS Conference on Human-Computer Interaction*, ser. HCI '10. Dundee, UK: ACM Press, Sep. 2010, pp. 98–107.

[36] M. D. Rugg, R. E. Mark, P. Walla, A. M. Schloerscheidt, C. S. Birch, and K. Allan, "Dissociation of the Neural Correlates of Implicit and Explicit Memory," *Nature*, vol. 392, no. 6676, pp. 595–598, Apr. 1998.

[37] D. L. Schacter and R. D. Badgaiyan, "Neuroimaging of Priming: New Perspectives on Implicit and Explicit Memory," *Current Directions in Psychological Science*, vol. 10, no. 1, pp. 1–4, Feb. 2001.

[38] S. Schechter, A. J. B. Brush, and S. Egelman, "It's No Secret. Measuring the Security and Reliability of Authentication via "Secret" Questions," in *IEEE Symposium on Security and Privacy*, ser. SP '09. San Jose, California, USA: IEEE, May 2009, pp. 375–390.

[39] S. Schechter, S. Egelman, and R. W. Reeder, "It's Not What You Know, But Who You Know: A Social Approach to Last-Resort Authentication," in *SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. Boston, USA: ACM Press, Apr. 2009, pp. 1983–1992.

[40] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[41] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Permier, "Oscillatory Gamma-Band (30-70 Hz) Activity Induced by a Visual Search Task in Humans," *The Journal of Neuroscience*, vol. 17, no. 2, pp. 722–734, Jan. 1997.

[42] N. B. Turk-Browne, D.-J. Yi, and M. M. Chun, "Linking Implicit and Explicit Memory: Common Encoding Factors and Shared Representations," *Neuron*, vol. 49, no. 6, pp. 917–927, Mar. 2006.

[43] D. Weinshall and S. Kirkpatrick, "Passwords You'll Never Forget, But Can't Recall," in *ACM SIGCHI Extended Abstracts on Human Factors in Computing Systems*, ser. CHI '04. Vienna, Austria: ACM, Apr. 2004, pp. 1399–1402.

[44] M. Wilson, "MRC Psycholinguistic Database: Machine-Usable Dictionary, Version 2.00," *Behavior Research Methods, Instruments, & Computers*, vol. 20, no. 1, pp. 6–10, Jan. 1988.

[45] M. Zviran and W. J. Haga, "A Comparison of Password Techniques for Multilevel Authentication Mechanisms," *The Computer Journal*, vol. 36, no. 3, pp. 227–237, Mar. 1993.

Appendix A
Details of the Pre-Study (Experiment 1) and the Long-term Study (Experiment 2)

	(9 days) 1st batch		(25 days) 2nd batch		(264 days) 3rd batch	
	No.	**Percent**	**No.**	**Percent**	**No.**	**Percent**
Age	97	100.0 %	129	100.0 %	124	100.0 %
20-30	61	62.9 %	66	51.2 %	69	55.6 %
31-40	27	27.8 %	40	31.0 %	38	30.6 %
41-49	6	6.2 %	14	10.9 %	12	9.7 %
50+	3	3.1 %	9	7.0 %	5	4.0 %
Gender	97	100.0 %	129	100.0 %	124	100.0 %
male	81	83.5 %	101	78.3 %	97	78.2 %
female	16	16.5 %	27	20.9 %	27	21.8 %
other	-	-	1	0.8 %	-	-
Country	97	100.0 %	129	100.0 %	124	100.0 %
France	40	41.2 %	55	42.6 %	54	43.5 %
Germany	41	42.3 %	44	34.1 %	45	36.3 %
other	16	16.5 %	30	23.3 %	25	20.2 %
Native English speaker	97	100.0 %	129	100.0 %	124	100.0 %
Yes	6	6.2 %	4	3.1 %	7	5.6 %
No	91	93.8 %	125	96.9 %	117	94.4 %
Profession	97	100.0 %	129	100.0 %	124	100.0 %
Administration	4	4.1 %	3	2.3 %	4	3.2 %
Arts	-	-	-	-	-	-
Engineering	38	39.2 %	55	42.6 %	52	41.9 %
Humanities	-	-	2	1.6 %	-	-
Life science	1	1.0 %	1	0.8 %	1	0.8 %
Science	53	54.6 %	66	51.2 %	65	52.4 %
other	1	1.0 %	2	1.6 %	2	1.6 %
Heard of Mooney images	97	100.0 %	129	100.0 %	124	100.0 %
Worked with before	-	-	-	-	-	-
Heard of before	10	10.3 %	11	8.5 %	8	6.5 %
none	87	89.7 %	118	91.5 %	116	93.5 %
Passwords are easy to remember	97	100.0 %	129	100.0 %	124	100.0 %
Strongly agree	4	4.1 %	3	2.3 %	4	3.2 %
Agree	30	30.9 %	32	24.8 %	36	29.0 %
Neither agree nor disagree	33	34.0 %	46	35.7 %	38	30.6 %
Disagree	27	27.8 %	39	30.2 %	39	31.5 %
Strongly disagree	3	3.1 %	9	7.0 %	7	5.6 %
Passwords are secure	97	100.0 %	129	100.0 %	124	100.0 %
Strongly agree	2	2.1 %	2	1.6 %	2	1.6 %
Agree	29	29.9 %	35	27.1 %	37	29.8 %
Neither agree nor disagree	36	37.1 %	38	29.5 %	37	29.8 %
Disagree	25	25.8 %	42	32.6 %	38	30.6 %
Strongly disagree	5	5.2 %	12	9.3 %	10	8.1 %
Mooney images are interesting to work with	96	100.0 %	129	100.0 %	124	100.0 %
Strongly agree	8	8.3 %	5	3.9 %	9	7.3 %
Agree	46	47.9 %	70	54.3 %	57	46.0 %
Neither agree nor disagree	36	37.5 %	39	30.2 %	48	38.7 %
Disagree	6	6.3 %	13	10.1 %	9	7.3 %
Strongly disagree	-	-	2	1.6 %	1	0.8 %
Using Mooney images is funny	97	100.0 %	129	100.0 %	124	100.0 %
Strongly agree	6	6.2 %	7	5.4 %	7	5.6 %
Agree	37	38.1 %	52	40.3 %	42	33.9 %
Neither agree nor disagree	40	41.2 %	56	43.4 %	59	47.6 %
Disagree	13	13.4 %	10	7.8 %	14	11.3 %
Strongly disagree	1	1.0 %	4	3.1 %	2	1.6 %

APPENDIX B
DETAILS OF THE FINAL STUDY (EXPERIMENT 3)

	(21 days) No.	Percent
Age	70	100.0 %
20-29	39	55.7 %
30-39	22	31.4 %
40-49	6	8.6 %
50-59	2	2.9 %
60+	1	1.4 %
Gender	70	100.0 %
male	54	77.1 %
female	15	21.4 %
other	1	1.4 %
Nationality	70	100.0 %
France	29	41.4 %
Germany	12	17.1 %
USA	9	12.9 %
other	20	28.6 %
Country you completing this in	70	100.0 %
France	37	52.9 %
USA	16	22.9 %
Germany	13	18.6 %
other	4	5.7 %
Native English speaker	70	100.0 %
Yes	10	14.3 %
No	60	85.7 %
Profession	70	100.0 %
Arts	2	2.9 %
Business	2	2.9 %
Engineering	24	34.3 %
Humanities	1	1.4 %
Life science	6	8.6 %
Science	35	50.0 %
other	-	-
Heard of Mooney images	70	100.0 %
Worked with before	3	4.3 %
Heard of before	18	25.7 %
none	49	70.0 %
Mooney images are easy to remember	70	100.0 %
Strongly agree	1	1.4 %
Agree	25	35.7 %
Neither agree nor disagree	35	50.0 %
Disagree	9	12.9 %
Strongly disagree	-	-