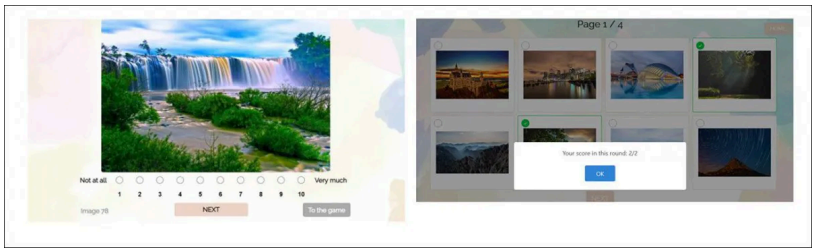
Aesthetic Preference Recognition as a Potential Authentication Factor

Originally published at https://www.unite.ai/aesthetic-preference-recognition-as-a-potential-authentication-factor/

A new paper from Israel has proposed an authentication scheme based on a user's aesthetic preferences, wherein the user calibrates the system one time by rating images, thereby generating a private 'domain' of that individual's visual and visual/conceptual predilections. Later, the user would be challenged at authentication time to match their recorded preferences against novel image sets.



From the trials of a 'game-ized' AEbA implementation – left, the user rates the aesthetic quality of an image; right, a score is signaled at the end of a stage in the a trials. Source: https://arxiv.org/ftp/arxiv/papers/2204/2204.05623.pdf

The system is titled *Aesthetic Evaluation-based Authentication* (AEbA), and is a submission to the 2022 USENIX Annual Technical Conference in California in July.

AEbA was trialed by the paper's researchers in the form of a game series, where participants were required to train the system and then rate new images that accorded with their registered tastes. A second round of tests examined a user's ability to guess the preferences of others.

From the paper – sample images, from pexels.com, suitable for usage in AEbA.

Such an approach may not be suitable for all people, since not everyone has a well-developed aesthetic sensibility, but could serve well either as a primary authentication scheme for low-medium security requirements, or as one choice in a range of possible adjunct methods in two-factor authentication (2FA).

However, the nascent idea of the system could form a starting point for more complex aesthetics-based challenge systems, since the number of images presented to users during authentication could be scaled up by default as necessary, in much the same way that CAPTCHA challenges can be prolonged in the event of uncertain initial results.

The more granular and extended the challenge, the higher the security such an approach can offer.

D	D_{HR}	*S*	Password space in bits	Memorized Secret authenticator
6	1	5	12	4 digits PIN
8	2	4	19	6-7 digits PIN
10	2	4	22	6-7 digits PIN
12	3	3	23	6-7 digits PIN
12	5	5	49	8 characters (72-character set)

A scale of relative password strength when several factors of an AEbA challenge multiply: 'D' represents the number of images displayed during the challenge; Dhr represents the number of images that the user is required to select; and 'S' is the number of screens (i.e. stages) in the linear process of aesthetic selection.

In terms of common conventions for human authentication, AEbA incorporates elements of *Something you know* (SYK) and *Something you are* (SYA)., and is predicated on three premises: that things we like (as represented in the visual realm) are easily distinguishable for us (in accordance with the general theory of mnemonics); our aesthetic tastes remain relatively consistent; and that there is adequate difference in the tastes of varied users to provide a non-guessable distinction in preferences.

The authors suggest that the technique could be adapted into machine learning frameworks capable of predicting individual users' evaluations.

The paper is titled *Beautiful secrets: using aesthetic images to authenticate users*, and comes from two researchers at the Software and Information Systems Engineering faculty at Ben-Gurion University of the Negev in Beersheba.

The Power of Image Domains

AEbA does not rely on memorization, but rather treats the end user as a trained image recognition system that has developed a robust and very specific gamut of pleasure responses, and keys in on these very strong pleasure associations.

In essence, AEbA hinges on the human equivalent of abstract priors in computer vision and image synthesis systems, which can convey style and domain-specific features without being embodied in a single and immutable image. It's through the application of such priors that a Generative Adversarial Network (GAN) can be trained to incorporate a domain (i.e. 'Van Gogh') into the generation of otherwise entirely novel pictures.

The new study posits evidence in prior literature that images are easier to memorize than words, that pleasing images are easier to memorize than general images, and that active evaluation of images (such as during the short AEbA training process) improves the memorability of images even further. Studies going back to the 1970s have established that humans possess 'massive storage capacity' for images in general, and for previously viewed images, and our ability to incorporate images into memory has been demonstrated to notably outstrip our capacity for verbal memory.

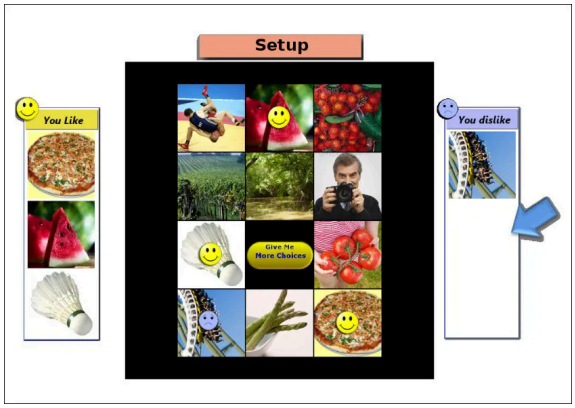
Though common sense suggests that domain experts, such as radiologists, would be most sensitive to images from their own domains, a 2010 study has asserted that memory capacity for everyday imagery is far more capacious than for domain-specific imagery, even in those with a visual 'specialty'.

Preference-Based Authentication

The notion of leveraging preference as an authentication mechanism came to prominence in two papers led by Markus Jakobsson of the Palo Alto Research Center, from 2008 onwards. This tranche of research around Preference-Based Authentication (PBA) suggested that music, food, artworks and other things that we like are ingrained in our minds and fueled by powerful internal motivations.

PBA was originally suggested merely as a device to facilitate password resets, using questions such as 'Do you like country music?', and concentrating on text-based preferences along traditional mnemonic principles, rather than visual input.

A subsequent collaboration from Jakobsson in 2012 substituted text with images:



A screen shot from the calibration/registration phase of the Markus Jakobsson 2012 PBA project. Source

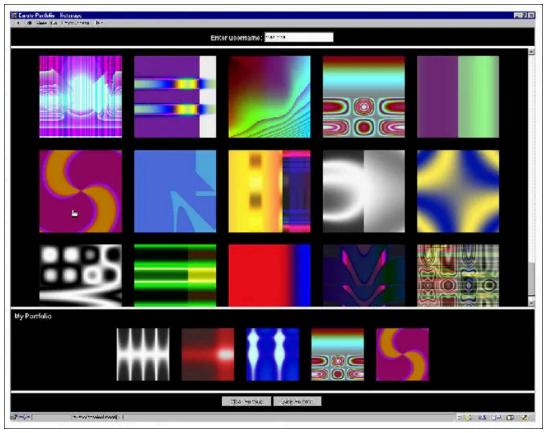
However, the authors note, this schema does not account for aesthetic evaluation of the images, but in effect uses pictures as proxies for words or concepts. By contrast, AEbA is seeking to discern a user-specific 'domain of pleasure' that's not directly related to specific things or activities.

The authors of the new paper also observe that there are practical limits to the number of items that can be presented to the viewer under the 2012 approach, whereas developing a more abstract model of user preferences removes these limits and makes external attacks and mimicry (i.e. based on phishing, personal knowledge, or other methods of subterfuge) far more difficult.

The idea of graphical passwords notably predates this work, with a proliferation of schemes emerging in the late 1990s. A contemporary study considers PassFaces, where users had to memorize faces (other than their own) rather than passwords. With this approach, a potential infiltrator would theoretically need an extraordinarily intimate domain knowledge of the user's facial preferences. Additionally, the user could presumably be relied on to select the same faces over time during the orientation phase.

From the late 1990s, the PassFaces scheme trialed at London's Goldsmiths University required the user to choose and memorize four faces of other people. The initial choice was based on the user's own preference, and in this sense the work is related to AEbA. Source

Most closely related to AEbA is *Déjà vu*, which presented viewers with random art images not necessarily designed to engage the pleasure response, but rather intending to use jarring and discordant imagery to help users memorize *specific images* that they would incorporate into a 'portfolio' during initial enrolment, and later be required to recognize from multiple possible images at authentication time.



Assembling a portfolio of 'preferred' images for Déjà vu. Source: https://netsec.ethz.ch/publications/papers/usenix.pdf

As the new paper's authors observe, this approach ignores the benefits outlined in neuroaesthetic literature (i.e. there is little internal motivation to connect with any possible images that are offered).

Additionally, such a method is vulnerable to 'shoulder-surfing', where a proximate (or MiTM) attacker may have an opportunity to witness which images are chosen. By contrast, a full implementation of AEbA would not repeat images previously used either in training or authentication sessions.

Additionally, the paper notes*:

> *'One of the problems identified in graphical passwords is that, like in regular passwords, users tend to select simple drawings, which decrease the variability of those passwords and make them more susceptible to adversarial attacks. Another problem (and perhaps a reason for the previous one) is potential interference if such schemes are used in several systems, i.e., users' memory of a password for one system impairs their memory of a password for another system. These issues are less of a concern when implementing AEbA, which relies on innate preferences that do not depend on specific accounts or on memorizing images.'*

The authors also emphasize an additional advantage of AEbA: contextual perception. Even if a shoulder-surfer or RAT attacker was able to view an authentication session, they would not know how far the 'unliked' images (i.e. presented images that the user rates lowly or rejects during authentication) are from the 'liked' image – a factor that will be different each time.

> *'Consequently, knowing that someone likes an image does not necessarily help if we do not know how much the image is liked relative to other images in the displayed set.'*

Additionally, it is impossible for a user to store their password insecurely for convenience, such as on a scrap of paper, because their domain of preferred image content is extraordinarily abstract and non-reductive.

Testing AEbA

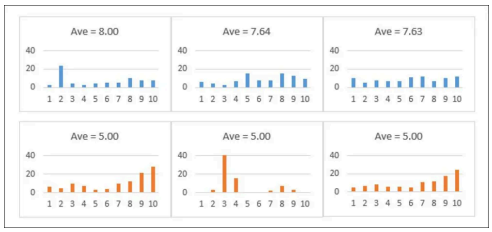
The researchers implemented the system as a game, in the context of a proof of concept of the project's core premises, curating a database of 318 images from free stock website pexels.com, and also including images from a personal archive.

The photos were classified into eight categories (*Universe*, *Nature*, *Mountains*, *Forest*, *Flowers*, *Cityscapes*, *Seaside*, and *Other*), and the trials divided into *Enrolment* (where the images were initially rated by the users in a one-off ten minute session), an *Authentication Game*, and finally an *Adversarial Game* (guessing the image preferences of others).

After weeding out non-contributing participants, the convenience sample (i.e. the trial group of participants) was reduced to 33 eligible players, consisting 21 females and 12 males.

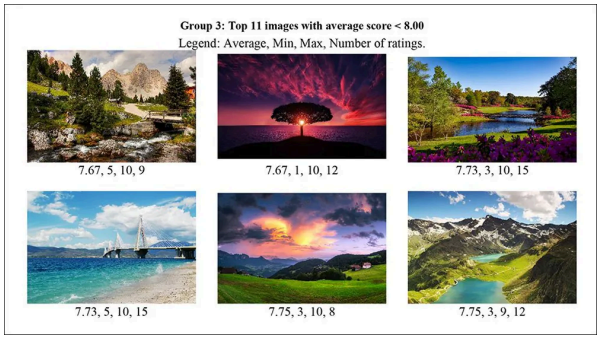
Enrolment

In the Enrolment phase, 3722 ratings were obtained for 274 images, with an average rating of 6.07, a median rating of 6, resulting in the most frequent values 7 and 8. The least-liked image scored just 2.32, and the most-liked 8.63.



The distribution of image ratings among top performers in the trials.

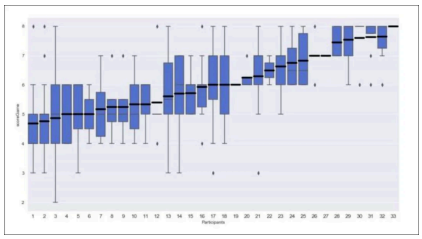
The authors contend that the notable skews towards high and low values in image rating, combined with the variety of such gradients across the user base, bears out their contention that users are able to apply highly differentiable liking scores to presented images, without the need to include obviously repulsive or 'out-of-distribution' images. It appears that the generally variegated whims and predilections across even a small user group are enough to validate the central concept.



Sample images with various user ratings.

Authentication

For the Authentication game, 264 playing sessions were conducted, with each participant completing the game twice over an average of eight sessions. Average success rate was 76%.



Box plot chart of game score distribution among the 33 members of the trial, with mean scores denoted in bold black horizontal line, displaying median, first and third quantiles, with minimum, maximum, and outliers.

Though there was a 'slight decline' in performance over time, this was greatly reduced among the top 50% of participants, practically disappearing in the 11 top participants (a third of the final user group).

Adversarial Game

The Adversarial Game component featured unrestricted play (unlike Enrolment), and occurred ten days after the launch of the Game phase. 190 games were counted for the results (excluding games where technical problems occurred). The average number of correct Adversarial choices came to 2.88, a 36% success rate technically equivalent to chance (particularly considering the low number of images in the dataset). However, in seven games, contributors were able to guess 75% or more of the correct images.

Conclusion

The casual test methodology (such as use of a convenience sample for testing candidates) in the study signifies that the approach currently represents a broad proof-of-concept; a nascent indication that human-centered 'domain capture' could one day provide an easy and even enjoyable method of authentication that is difficult to appropriate or interfere with. It's clear that much more rigorous trials, with higher numbers of participants and a properly-staged authentication scenario would be needed to establish the value of AEbA.

The authors conclude:

> *'It would also be interesting to study the potential of using machine learning techniques to predict individual users' evaluations and to generate keys and decoys that the user has not previously rated. Doing so could increase the password space by increasing individual users' image pools and their variability.'*

**My conversion of the authors' inline citations to hyperlinks*

First published 13th April 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai