Visual Secrets : A recognition-based security primitive and its use for boardroom voting

Enka Blanchard
Digitrust, Loria, Université de Lorraine

Sébastien Bouchard
LaBRI Université de Bordeaux

Ted Selker
Cyber Defense Lab (UMBC) Selker Design Research

Abstract

This paper presents and evaluates a new security primitive in the form of non-transferable "visual secrets", and an application at the center of a low-tech visually verifiable boardroom voting system. Visual secrets rely on the pre-semantic treatment of images in the human brain. After being shown an image for a limited time, users can recognise it when mixed in a larger set, but cannot reliably communicate to someone else exactly how to do so — whether voluntarily or through coercion.

We report on a usability study on 151 subjects which showed that they could recognise an image they had previously seen when shown among 20 similar images with an accuracy of at least 79% compared with an expected baseline of 5%. Despite their recognisability, the "secret" images were hard to describe in unambiguous ways : no assessor managed to accurately identify the images from the description given by the subjects.

We then introduce a boardroom voting system based on this primitive. The voter receives a ballot consisting of a single picture, votes by folding it horizontally or vertically and casts it. When all ballots are revealed, the voter can check with a glance that their ballot is present and folded correctly. This gives them the opportunity to detect error or fraud without being able to reveal to others how they voted. The design makes use of textured paper to provide both accessibility for the blind and improved usability for all users.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Vancouver, B.C., Canada.

1 Introduction : defining visual secrets

Researchers in usable security often talk about "something you are, know or have". Those secrets are often shareable : one can give their home keys to a friend, be coerced into revealing passwords, or even have their biometrics such as fingerprints stolen [21]. Is it then possible for humans to have useful secrets that cannot be shared ?

Let us suppose two individuals decide to meet in public and want to be able to ascertain each other's identity. However, they are afraid of one of them being coerced into revealing the identification mechanism, and being replaced by an adversary. Any passphrase or callsign could be obtained under coercion and replicated. The problem then is to find a secret they would recognise but would not be able to share, no matter the context.

In a formalised version, this problem is not *a priori* solvable by independent agents in a classical computing setting. As Turing Machines can simulate each other, any communication between agents would be indistinguishable if one agent were simulated. This is not necessarily true in a quantum setting, as a non-simulatable protocol could potentially be found, thanks to the no-cloning theorem — depending on the formalism used [31].

However, humans are not Turing Machines, and they have unique abilities. One option is to use something that only they could do, for example a behavioural biometric. This is possible in the abstract case, but multiple problems exist with those, from high error rates to biometric identity theft [8, 25]. Moreover, this type of secret can require complex apparatus to measure.

A second lead is then to use specialised human cognitive functions. Multiple advances have been made in this direction over the last decade, mostly in the context of authentication [10, 22]. There have been some measure of success in creating unshareable secrets in [6], as subjects have no conscious recollection of them, but the training is time-intensive (at least 30 minutes for one password). One cognitive function of particular interest to us is linked to image recognition. As has

been demonstrated since the 1960s, humans have an extensive memory for visual stimuli [18, 23, 32, 34]. This has already been used as a source of security primitives, for example with authentication in the case of visual passwords [15], as well as with various biometric methods [2, 36]. Most importantly in our case, a significant aspect of this image recognition happens in a pre-semantic and pre-cognitive fashion, requiring no conscious effort, thanks to specialised neural pathways in multiple areas of the brain [18, 26]. This is related to the difference between recognition and recall [14]. The mind's pre-semantic treatment means that there might be a loss of information during image recognition. As such, the ability to recognise an image is not directly related to our mental description of it, and any description might ignore some key elements of the picture[1].

The approach takes inspiration from both this cognitive science research and concepts from zero-knowledge proofs [12]. This pre-semantic treatment is used as a source of secrets that are recognisable but not shareable, and we call the resulting primitive a *visual secret*[2]. A user with unlimited time and good eyesight might be able to describe exhaustively each pixel of an image. However, practical protocols would have reasonable constraints on the time spent describing images.

These constraints are especially appropriate in our case, as the first proposed application of visual secrets concerns verifiable voting. We propose a low-tech solution to the problem of *boardroom voting*. This corresponds to a small group of participants — e.g., jury members — having to quickly vote on an issue, generally between two possibilities. In practice, such votes are often held informally by writing an answer on a piece of paper, although a variety of electronic and low-tech options have been proposed [1, 4, 13, 24]. The central idea behind our application is to have a visual secret on each ballot and to count them publicly. This allows each voter to check that their ballot is present and counted correctly, but prevents them from proving to someone else that they voted a certain way. For this application, two metrics are crucial : a high short-term recognisability (to find one's own ballot) and a low describability (to prevent sharing how one voted).

The first part of this paper explores the viability of this new approach with a focus on these metrics. We start with a description of the empirical study and an analysis of its results. We then introduce the voting protocol, discuss the results, and conclude.

Main results

This paper features three main contributions :

- visual secrets, a new security primitive ;
- the results of a usability study on 151 subjects that demonstrates that visual secrets have both high recognisability and low describability ;
- a first application of visual secrets in the form of a low-tech verifiable voting system.

[1] This principle is already used in police lineups, in which multiple suspects corresponding to the description are shown and where the witness is supposed to be able to find the one they saw previously.

[2] Visual secrets are not related to visual cryptography [27, 28].

2 Empirical study

The goal of the study was to test the viability of visual secrets as a security primitive. Subjects were shown pictures and had to describe them, before having to find their initial pictures among a larger set. As we conjectured that the recognisability and describability of the pictures would depend on what they depict, three different image series (lions, mountains, and abstract shapes) were included. The study had three main objectives :

- test the ability of subjects to find their picture (after spending a few minutes on other tasks) (*recognisability*) ;
- test the ability to describe their picture accurately and unambiguously (describability) ;
- if possible, compare the three image series on the previous two metrics.

The hope was to find at least one image set with high recognisability but a low describability.

2.1 Protocol

The online protocol was split into four sections (illustrated on figure 1) :

1. A single introductory page informing subjects of their rights (including the right to quit at any point) and informing them that they would have to confirm at the end to submit the experimental data. It also asked whether they had performed or seen someone else perform the experiment and whether they were on a mobile device.
2. Three pages, with each featuring a picture (one per series). The instructions were to describe the picture in at most 10 words to try to make it identifiable among similar images.
3. Three pairs of pages, with each page featuring 2 rows of 5 images. Each pair of pages corresponds to a series of 20 images. These images were randomly distributed between the two pages, in such a way that all images were shown exactly once. Subjects were asked to select an image if they thought it was one they had seen earlier, and could also select "none".
4. A conclusion page thanking them for their input, indicating their scores on the memory phase, and asking them to confirm the submission of the experimental data.

A/B testing was used to randomly assign the order of the first two image series (lion and mountain), with the image recognition order being the same as the presentation order. The third series was always shown last. The tasks of writing the second and third images descriptions then served as distractor tasks in order to limit the effect of short-term memory. The A/B testing allowed us to measure and compensate the effects of order and delay — on recognition (which was not statistically significant).
The experiment was tested with an informal pilot study[3] among colleagues before being put online at *redacted for anonymity*.

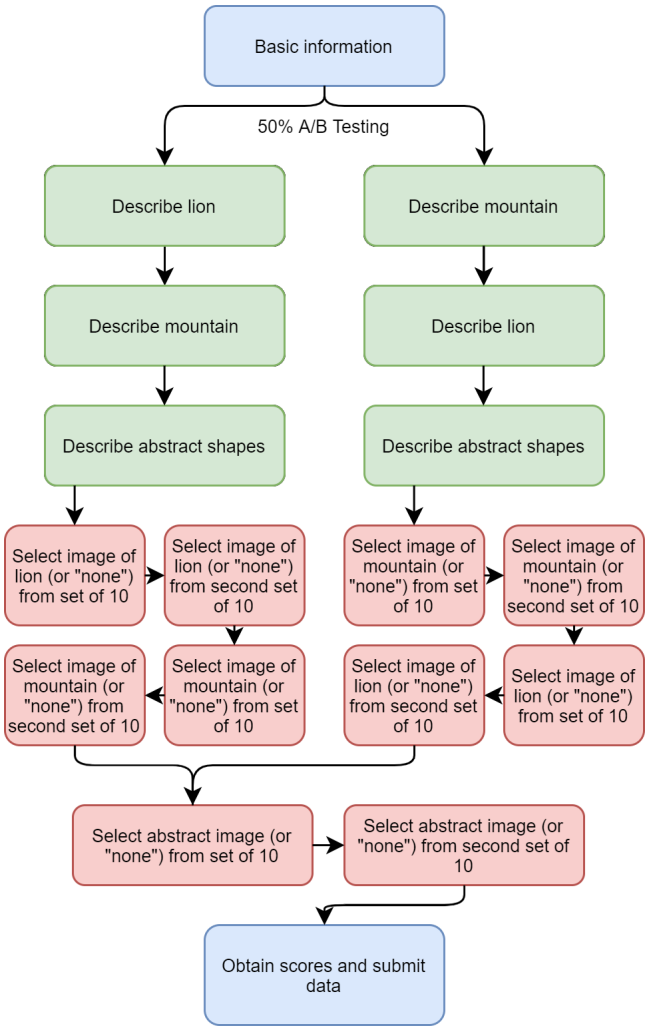

Figure 1: Diagram of the experimental protocol

[3]The pilot study data is not included in the analyses as the protocols differ slightly.

2.2 Measurements

The only two questions not directly relevant to the study were whether the subjects used a mobile device (as it changed the image layout), and whether they had participated or seen someone participate in the study (as it could affect the memorisation). Considering the experiment's statistical power, we did not expect to be able to distinguish differences in demographic base performance. Thus, and out of a general concern over studies featuring irrelevant demographics questions, we decided to collect as little identifying data as possible — in accordance with local legislation.

All other recorded data relates to the answers provided to the questions asked during the study. For each of the three series of pictures, we recorded :

- the index of the picture assigned to the subject (1 to 20) ;
- the description they gave for it ;
- the list and order of each of the two sets of 10 images shown ;
- the indices of the pictures they recognised, with a zero indicating that they chose "none" ;
- how much time they spent on each page.

We also recorded which A/B testing group they were in.

2.3 Image bank

Three series of 20 pictures with three themes : lions, mountains, and abstract shapes were selected for the experiment. The pictures were either (free of rights) pictures of lions or mountains, or they were abstract shapes randomly generated by the authors. The three image series are shown in Figure 2. Section 6.1 discusses our decision not to include human faces.

2.4 Subjects

Data was collected from September 1st, 2020 to December 31st, 2020. Subjects were recruited through John Krantz's Psychological Research on the Net index [19]. A total of 164 subjects participated in the study. All but two of them wrote their answers in English (with one French and one Spanish). The median time spent on the experiment was 213 seconds with a standard deviation of 169s — discounting users who took a noticeable break (between 20 minutes and 15 hours).

Any subject who had not provided intelligible answers to the first part of the protocol was removed from the dataset. This eliminated a total of 13 answers, mostly corresponding to subjects who had skipped the questions, as well as a few who wrote descriptions such as "pee" or "po".This removal is not targeted towards the worst-performing subjects : of the 13 removed, 6 actually had perfect memorisation scores.

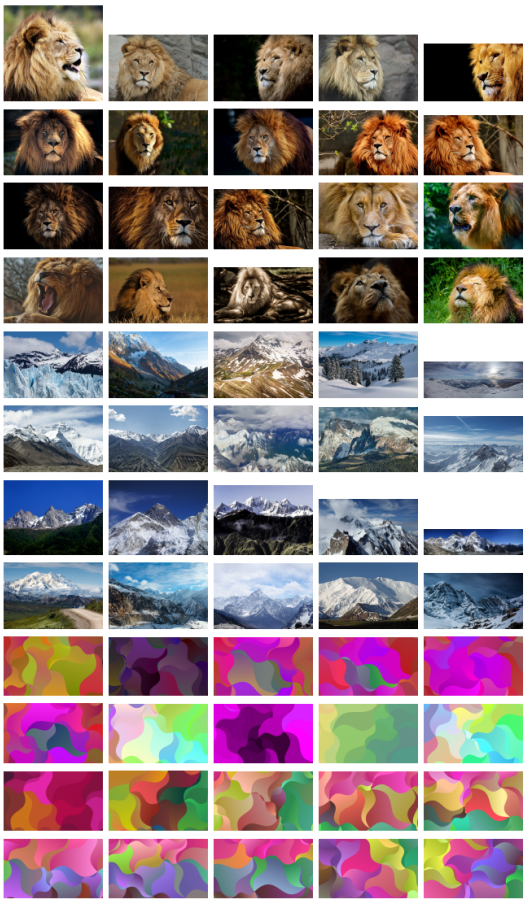

Figure 2: The three image series used for the experiment.

A total of 4 subjects indicated that they had seen someone perform the experiment before or had performed it themselves. We chose to include them as their performances are similar to the other subjects and as such have little impact on the general results[4].

3 Empirical results

This section features the analyses performed on the data collected, in the goals of estimating the recognisability and the describability of the different image series. In the goal of transparency, all the data in this section has already been put online in a publicly accessible archive at : *redacted for anonymity*.

Before detailing the results, we must make a note on statistical analyses. We were hoping to observe a difference between the human recognition error rate and a randomised algorithm — a null hypothesis corresponding to a 5% success rate.

However, we were mostly trying to observe the magnitude of the effect. To this end, we have more than enough statistical power as the empirical data (>79% success rates) corresponds to extremely high values (z-scores >40 for all series, corresponding to p-values $< 10^{-350}$). This establishes statistically that visual secrets are recognisable. However, because the error rates are all lower than expected, comparing them between the series is not within the statistical power of the experiment, and the differences reported in the data should be taken as means that are not statistically differentiable from one another due to high variance (hence, no comparisons are made between the series as they would all have p>0.05). Assuming the means are accurate, acquiring statistical power for a rigorous analysis of variance between the image series would have required more than 1 000 subjects.

3.1 Error types and frequencies

The analyses in this section are split by series (lion, mountain, and abstract shapes), merging the two groups used in the A/B testing as an analysis of those showed no difference at all.

We interpret and analyse the answers to the image memorisation tests as a two-answer test rather than two independent tests. The subject succeeds if they get both answers right : if they manage to both find the image that had been shown earlier and click “none” on the other test. This allows us to compare their performance with an equivalent memory-less algorithm which chooses an image when not primed with one. Without the two-answer approach, such a memory-less algorithm on independent tests would skew the results as the optimal strategy would be to always pick the “none” option, which is neither useful nor realistic. On two-answer tests, the memory-less algorithm would have at best a 5% success rate[5] (corresponding to picking one image from each set of 20).

The following categorisation of errors was applied to each two-answer test :

- **False negative** (FN) : the subject answered “none” to both queries ;
- **False positive** (FP) : the subject correctly found their picture but also recognised a second picture ;
- **Single mistake** (SM) : the subject correctly answered “none” to one query but chose the wrong image on the other query ;
- **Complex mistake** (CM) : the subject had a false positive on one query and a false negative on the other ;
- **Double mistake** (DM) : the subject chose the wrong image on one query and had a false positive on the other query.

[4] Another factor for including them is that it is not evident in which direction the results would be affected if we excluded them : repetition could improve the performance, but could also increase the rate of false positive recognition.

[5] This corresponds to an optimised memory-less algorithm, with a naive one having an 0.8% success rate.

Thus, a subject either accurately finds their picture ('success') or commits one of the previous four kinds of errors. The main subject accuracy performances are indicated in Table 1 :

	Correct	FP	FN	SM+CM+DM
Lion	125 (83%)	12	6	8
Mountain	130 (86%)	8	10	3
Abstract	120 (79%)	12	9	10
1st image	128 (85%)	9	9	5
2nd image	127 (84%)	11	7	6

Table 1: Subject accuracy and error types for each series as well as for the first/second image shown (abstract always being third).

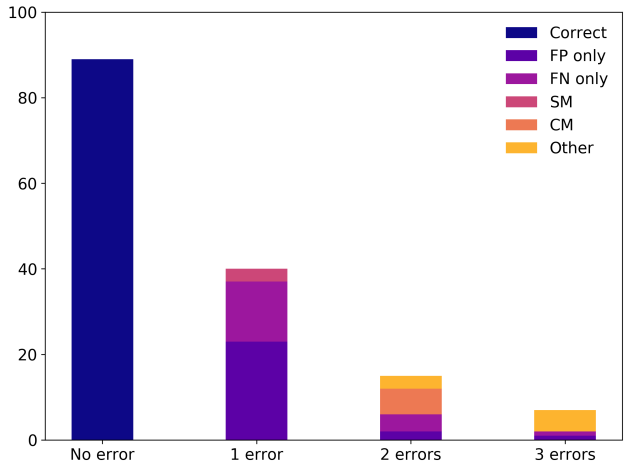

Figure 3: Number and types of errors made by the subjects for the memorisation questions.

Figure 3 above shows the error types and the number of errors when trying to find a previously seen picture. Among the 7 people who made three mistakes, one had only false positives and one had only false negatives. This might be due to misinterpreting the instructions, an effect also observed and confirmed when running the pilot version of this experiment among colleagues.

We computed the error rate independently for each image, which gave no statistical result. As could be expected, all the images have statistically similar rates of being chosen incorrectly, and the ones with a higher proportional error rate correspond to the ones that were shown the least[6] — thus increasing the proportional variance. We also looked at whether the recognition errors tend to form clusters of related images, but the corresponding graphs are too sparse to hypothesise the presence of any definite cluster.

[6]In practice, no image was misidentified more than 3 times.

3.2 Image describability

To estimate image describability while minimising bias, two of the authors independently categorised the full list of descriptions subjects wrote about their assigned images, before comparing with the rest of the data. For each description, the assessors selected all images that could potentially fit — without knowing what the correct answer was. One of the assessors had the instruction to be strict in their estimates, and the other had the instruction to be lenient[7]. This provided upper and lower bounds for multiple metrics.

As can be seen on Figures 4-6, some descriptions provide no information at all as they potentially correspond to all images. This is not always related to concision : although "lion" was the full description given by multiple subjects, some also provided other non-distinguishing descriptions like "the lion is the king of the jungle" or "70s, groovy, Brady Bunch, swirly, woman, colorful" (for the abstract image). We call these descriptions *trivial* if they can correspond to any image — for a given assessor. We discount them in some analyses to have a more rigorous interpretation of describability[8]. A few descriptions resulted in very different assessments, such as "the lion is focused on something", where the lenient assessor selected 17 images whereas the strict assessor selected no images.

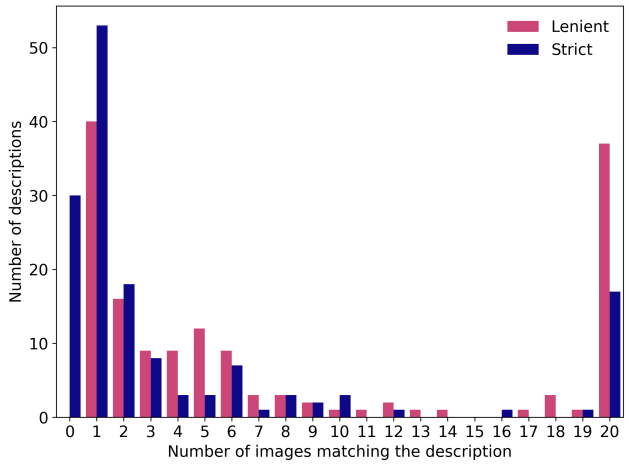

Figure 4: Distribution of the number of selected images by the strict and the lenient assessor for the lion series.

[7]In multiple cases, this meant that some parts of descriptions were ignored as potentially small mistakes. For example, 18% of subjects describe the abstract image as having blue among its main colours, although blue is very rare in the image set, and only twice makes up more than 15% of the image (when including many shades of blue).

[8]Our hypothesis was that most users would provide ambiguous descriptions showing the difficulties in sharing their secret. Eliminating the worst performers imposes a stricter threshold on any subsequent result.

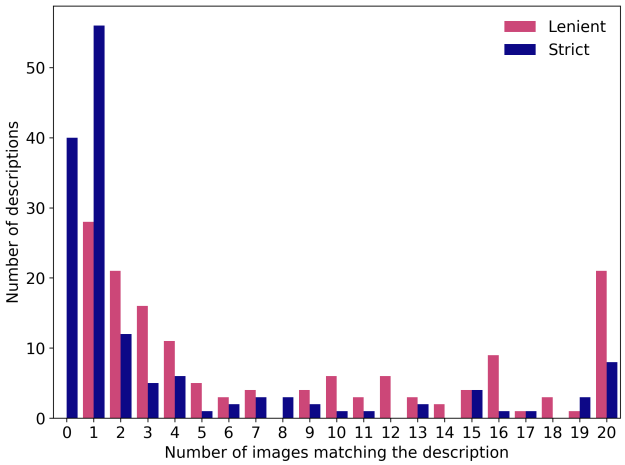

Figure 5: Distribution of the number of selected images by the strict and the lenient assessor for the mountain series.

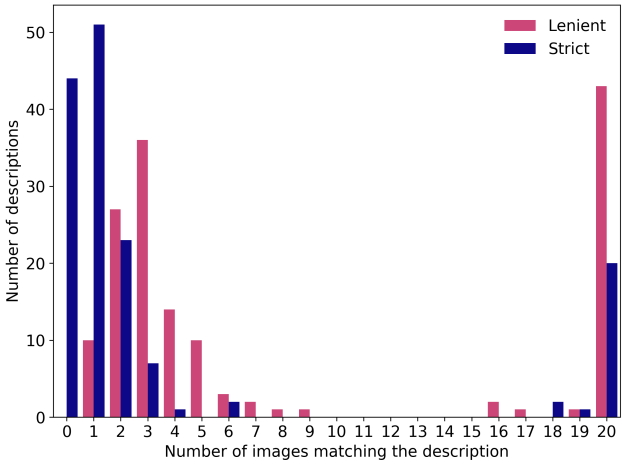

Figure 6: Distribution of the number of selected images by the strict and the lenient assessor for the abstract series.

	Assessor	Lion	Mountain	Abstract
All descriptions	Strict	59% (89)	55% (83)	50% (76)
	Lenient	85% (129)	78% (118)	76% (115)
Non-trivial descriptions	Strict	54% (72)	52% (75)	43% (56)
	Lenient	81% (92)	75% (97)	67% (72)

Table 2: Accuracy of the categorisations by the strict and the lenient assessors for the three series of images considering either all descriptions or only non-trivial ones (percentages with total number in parentheses).

The accuracy of both categorisations (as the fraction of image selections that included the originally described image) is shown in Table 2. In the goal of assessing the security of the images as potential visual secrets, one question is crucial : can they be accurately and unambiguously described, or in other words, does a description fits a single image ? Even with this noisy dataset, there is some evidence that certain unique elements get picked up by most subjects. Table 3 shows for each image series and assessor the number of unambiguous descriptions. It also shows how many of those descriptions considered unambiguous were in fact attributed to the wrong image (a low accuracy making the system more secure).

	Assessor	Lion	Mountain	Abstract
Correctly unambiguous	Strict	36	40	35
	Lenient	32	23	7
Wrongly unambiguous	Strict	17	16	16
	Lenient	8	5	3
Unambiguous accuracy	Strict	68%	71%	69%
	Lenient	80%	82%	70%

Table 3: Number of unambiguous identifications by the strict and the lenient assessors for the three series of images. The proportion of correct identifications among unambiguous images is shown on the bottom lines.

Finally, we tried to compute clusters of images likely to be all selected by either assessor when one of them is described, therefore indicating images for which it is unlikely that the subjects provide a description distinguishing one from another. A first step was to find whether certain images often get confusing descriptions. For example, when faced with descriptions of abstract image #2, the lenient assessor considered that image #6 was potentially the one being described twice as frequently as image #2. Table 4 shows the number of images for which the descriptions generally point to a different image.

	Most probable		Among probable		Not probable	
	Lenient	Strict	Lenient	Strict	Lenient	Strict
Lion	10	12	6	3	4	5
Mountain	8	10	6	4	6	6
Abstract	6	8	6	5	8	7

Table 4: For each series, this table shows the number of images for which the descriptions tended to correspond to other images more than their original image. For each image, if its descriptions most frequently point to itself being selected; it is counted in “Most probable”. It is counted in “Among probable” if there are other images as frequently assigned to its descriptions, and “Not probable” otherwise.

We also computed a full graph for each (assessor, series) pair. Despite the limited number of descriptions, none of the graphs are sparse (with 145 to 304 edges out of a maximum of 484), making them hard to interpret. Figure 7 shows one such graph (the most legible one, having the least number of edges). The noisy nature of the dataset limits the interest in deleting the low-weight edges.

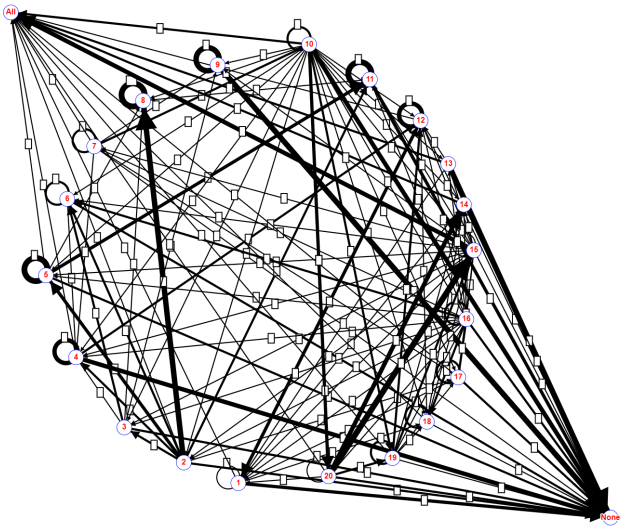

Figure 7: Graph of the abstract shape descriptions by the strict assessor. Node size is proportional to how often it was described, and edge thickness to the how often a description for i was interpreted as one for j, with two added nodes ("None" and "All") to merge trivial descriptions and make the graph more legible.

Despite the difficulties in handling this noisy quantitative data, some effects are apparent, and confirmed by qualitative feedback. For example, the lenient assessor managed to unambiguously recognise every description of the #20 lion picture (and the strict assessor performed almost as well). The peculiarity is that the unique features pointed out by the subjects varied a lot, e.g. the sleeping/closed-eyed lion or the yellow flowers. This image was not identified as particularly describable when creating the dataset, so care should be taken when creating new datasets as empirical validation is crucial.

4 Visually Verified Ballots (VVB)

We now describe a first application of visual secrets in the form of a low-tech voting system appropriate for boardroom elections.

4.1 Ballot design

The ballots look and feel like square cards. Just like cards, one side is left blank — or with a regular symmetrical pattern — and the other has the relevant information. The visual information is minimal, as it consists of two elements :

- A picture from a common set of visual secrets, covering the whole card ;
- Two orthogonal lines crossing the picture, respectively labelled with “Vote 1” and “Vote 2”.

This visual information is complemented by tactile information in the form of texture — bumps — present on both ends of each line. This texture corresponds to International Braille for A/B, with one bump for the first line and two for the second line[9]. Care should be taken when applying the tactile patterns[10] to avoid the bumps being noticeable from the other side — as well as to avoid transparency issues. Moreover, ballot should ideally be thinner along the horizontal and vertical lines to make folding easier. An example of one such ballot is shown on Figure 8.

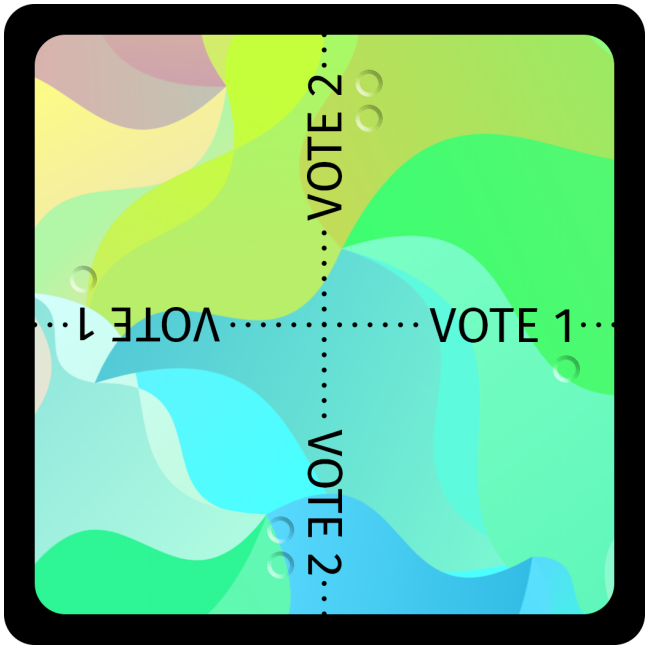

Figure 8: Example of a Visually Verified Ballot. Folding along a line correspond to voting for that option. The embossed bumps represented under the folding lines help voters keep track of which line is which.

Visually Verified Ballots should be made available as packs of 20 to 50 ballots, wrapped and sealed like a playing card deck[11].

4.2 Protocol

The protocol goes as follows :

1. The vote organiser opens a new pack of ballots in front of all voters ;
2. One ballot is distributed face down to each voter ;

[9] The fact that it corresponds to Braille is only of limited benefit, as less than 10% of blind children in the USA are taught to read Braille [30].

[10] For example, using thick ink instead of mechanical embossing.

[11] Initial estimates show that the cost to manufacture such packs should be between $0.50 and $1 per pack, using commercial printing services.

3. Each voter lifts up their ballot to privately look at the image and memorise it ;

4. Each voter rotates their ballot a few times, while keeping track of its orientation using the bumps ;

5. Each voter folds their ballot along the line of their choice to select “Vote 1” or “Vote 2” to be on the fold ;

6. The voters cast their ballots in a ballot box or a bag ;

7. The ballot box is upturned and all the ballots are unfolded on a table in front of all the voters’ eyes ;

8. The vote organiser tallies the votes orally while the voters check that the ballot featuring their assigned picture are present with the correct fold ;

9. The vote organiser announces the result and the vote is over unless someone challenges the result.

5 Security considerations on VVB

Our adversarial model relies on the fact that the context is an in-person boardroom election. This means that the cost of being caught is high and that the first priority of the adversary should be to remain covert and as unsuspected as possible. If caught, they could be banned and the election would proceed without them with little delay, as opposed to a large-scale election. As done in [4], we assume that the adversary can have some accomplices, as well as some skills in prestidigitation, allowing them to dissimulate and manipulate paper ballots in a discreet way. However, we assume that they cannot convince or coerce the majority of voters into doing their bidding[12]. The adversary can have any of the following objectives :

1. directly change the outcome of the election ;

2. find out how other people voted ;

3. coerce others into voting for a designated candidate ;

4. cast doubt on the outcome (discreditation attack).

Abstract visual secrets are used for the analyses, as they are the worst-performing in terms of recognisability and thus give a lower-bound on the expected performance of the other image series (as small variations in describability have limited impact).

[12] As many elections use a simple majority, this would make many objectives irrelevant. It also opens multiple new avenues of attack depending on the degree of control the adversary has on the other voters.

5.1 Changing the outcome

Assuming the ballots are publicly counted, an adversary wanting to change the outcome must manage to do at least one of the followings :

- add ballots ;
- remove ballots ;
- replace ballots.

As there are at most a few dozen voters and the vote counting is done in plain view, discreetly adding or removing ballots is not feasible. The goal is then to replace ballots without voters noticing.

If all voters could recognise their ballot perfectly, there would be only one way to achieve this : by having one ballot correspond to multiple voters. This would not be unfeasible but would require a complex attack with a specially made pack of visual secrets, the ability to distribute those selectively, then to remove them from the ballot box before adding other ballots.

Let’s then consider that voters do not have perfect recall, but instead have a performance similar to the data from Section 3.1. Then changing the fold on a single ballot would get noticed with at least 79% probability[13]. Changing multiple ballots without anyone noticing would then have an exponentially small probability (in the number of modified ballots).

However, there might also be false positives. Assume that there are twenty voters with a split vote[14]. We would then expect 2 or 3 voters to be mistaken or confused about their ballots (but probably not sure of themselves). Having 3 voters notice errors (and be sure of themselves) would then be a strong indicator of malfeasance, and would be a probable outcome even if only a few ballots are replaced.

5.2 Finding out how other people voted

There are four ways to find out how someone voted.

First, one could keep track of which visual secret got to which voter during distribution — if all the images are known beforehand. This would require having two identical unopened packs of visual secrets, and for the adversary to be able to impose using one of those packs. Moreover, they should then control the distribution of the pack of visual secrets. Having someone shuffle the pack beforehand would then address this issue as long as the person shuffling is not an accomplice. Attacks would still be possible, but would require technical ability, equipment, and the presence of at least one accomplice.

[13] However, folding a ballot twice would in itself leave marks).

[14] If the vote was strongly in favour of the adversary’s choice, there would be no need to cheat, and if was strongly in their disfavour, the following analyses would give even higher chances of getting caught).

Second, one could make an identifying mark on the back of certain ballots, observe which voters got them, and then look for those marks during the count. The difficulty lies in the fact that all the ballots are observed by all voters. A mark visible from a distance would have a high chance of being caught during the counting. A related risk would be for voters to fold ballots incorrectly in a noticeable fashion, but the folding lines and the number of voters limit this attack's feasibility.

Third, one could have another voter discreetly show their ballot during the voting period (or take a picture with their phone), as proof that this is their visual secret. However, all the voters are under high scrutiny as everyone can see each other. They should only have time to get their ballot, think briefly, rotate and fold it, and then cast it. As such, showing one's ballot to someone else (except maybe one's immediate neighbour) would be risky and the culprit could get caught.

Finally, one could ask another voter to describe their ballot. Once the ballots are public, nothing prevents the voter from lying and describing a different ballot. If they describe their ballot before the count, they must do so discreetly and succinctly as they are in the same room with limited time — also reducing the scalability of such process. However, even in the best case, no assessor managed to identify accurately and without ambiguity more than 26% of a series of visual secrets from their descriptions[15], limiting the interest of such a method. Moreover, the subjects were describing their pictures while being able to observe them, as opposed to describing from memory, which would further reduce performance.

5.3 Coercing others

The most common way of coercing people into voting for a designated option is to have a way to find out how they voted. As per the previous subsection, that should not be doable except for the voters immediately next to the adversary.

The other standard option is to vote in someone's stead and prevent them from reporting it — or to give them an already filled-out ballot, as with chain voting [17]. Voting directly in someone else's stead is impossible in a boardroom. However, if a bag or an opaque box is used as a ballot box, one attack could hypothetically be achieved. It would require the coerced voter to simulate dropping their ballot in the bag (while keeping it in their hand). The adversary could then add two ballots simultaneously without alerting anyone. The difficulty with this attack is that it requires some sleight-of-hand ability in both the adversary and the coerced voter, and is irrelevant if the ballot box is transparent.

If they can't check how people voted or force them to vote for a given candidate, the coercer cannot enforce their will.

[15]The assessors were categorising the descriptions independently, as the same visual secret could correspond to multiple descriptions. Having the constraint that there be a matching between the set of secrets and the set of descriptions would improve the assessor's performance, but the constraints would probably still be too weak for the method to be reliable.

5.4 Discrediting the election

A small group of adversaries could falsely claim that their ballots were modified to create a discreditation attack — presumably cancelling the vote. This is the case with all verifiable voting systems in which voters can only detect but not prove the existence of malfeasance. As such, our system is vulnerable to this kind of attack. However, as all voters are in the same room, they can easily start a new election, potentially changing the people in charge of handling the election or even using a different system — ideally more secure. This kind of attack would then mostly delay the election by a small amount while bringing unwanted attention to the group of adversaries. Moreover, if the election is to be decided by a simple majority and a single voter reports irregularities while the margin is of at least a few votes, the voters could decide (in advance) to accept the result.

6 Discussion

6.1 Design choices

A central design choice was in the theme selection for the image series. While people are in general much better at recognising faces than any other image, we chose to avoid human faces in this experiment for the following reasons :

- Most humans[16] have specialised neural pathways that react specifically to faces [16, 29], which could create unrepresentative stronger reactions than can be expected for other images.
- Specialised facial processing's performance depends on the age, ethnicity, and gender of the face shown to the subject [39, 40] .
- Languages tend to have specialised vocabulary to describe faces — which is more widely spread than the technical vocabulary required to describe a mountain — which could improve the performance on describability.

For these reasons, we chose to use non-primate faces as they would not trigger human-specific responses [11]. As we still wanted to compare different types of images for describability, we settled on animal faces (lions) and natural scenes (mountains) [7]. For the third series, we wanted to have abstract images as they have the advantage of being easy to generate automatically — and as we conjectured that they would be harder to recognise and especially to describe. We restricted the study to these three series to limit the time spent by subjects and the drop-out rate.

[16]This sets aside people with certain neurodivergences and ones suffering from prosopagnosia, representing a non-negligible subset of the population [37, 38].

6.2 Limitations of the study

This study has one main limitation. It tested the subjects' memory only a few minutes after the initial stimulus. Although writing the other descriptions in between provided a distractor task, the recognisability might still be influenced by short-term memory effects. The impact of time spent between the memorisation and the recognition had no evident (or statistically significant) effect. There was a tendency to have slightly lower recognition when considering subjects who had a delay between 3 and 15 minutes, counteracted by the fact that many of the slowest subjects had perfect scores.

A second limitation is that the subjects were recruited from one of the major sites that indexes psychological studies online (with around 500 studies hosted each year), which could have created a recruitment bias. However, previous studies have shown that the participants from this subject pool tend to give higher-quality data than users of Mechanical Turk, and that they cover a wide range of demographics, albeit with a bias in favour of college-age respondents [20].

A third potential limitation lies in the study's design as a web experiment where data is only stored if the user confirms at the end. First, this could limit the ecological validity compared to physically interacting with the pictures in a laboratory experiment with a controlled environment. Second, we could not measure the drop-out rate, and, more importantly, the proportion of subjects who dropped out and redid the experiment. Two factors mitigate this. First, only a few subjects indicated having performed the experiment or seen it performed earlier. Second, other studies using the same source of subjects recorded a limited drop-out rate and next to no repeating subjects [5].

6.3 Considerations on using VVB in practice

From what we have shown, VVB should be reasonably secure when used with small groups of voters who can see each other and who vote rapidly, thus benefiting from the potential short-term memory advantage. One application that directly comes to mind is for juries in legal cases. VVB is particularly suited to this situation for multiple reasons : the set of voters is small and they do not know or trust each other ; they might want to organise multiple votes, so each one should take little time ; they are not expected to be skilled at sleight-of-hand. As juries are often decided shortly before a case, it is also hard to train a jury member to perform certain attacks.

One limitation of VVB is that the version shown above does not allow voters to anonymously abstain but only to vote between two possibilities. One variant shown in the next subsection addresses this by extending the ballot to more than two candidates.

Another limitation of VVB is in its accessibility to blind voters. As it is designed, blind voters can vote with privacy and without assistance — unlike in many other voting systems. That said, VVB doesn't let blind voters verify their ballots[17]. This is a limitation of the system, but the security of blind voters is still assured by the fact that their ballots are not identifiable. That said, if blind voters are present, care should be taken during the counting to check that the texture information corresponds to the written information.

We should also warn that as it is, VVB should not be used for elections with more than a few dozen voters. The ability to find one's visual secret has only been tested among 20 images, and not among a set of 200. It might be possible to generalise the method to handle more voters, but the viability of visual secrets in this context has not been tested.

6.4 Refinements on VVB and variants

A first type of modification would be to address the fact that the VVB shown here only allow binary elections. They are easily adaptable, however, by using circular ballots or polygonal ballots with $2n$ sides and n folding lines — one per candidate — inspired by what was proposed in [4].

There is also the question of the order of candidates on the ballot, which generally has a non-trivial impact[18]. Thanks to the symmetry and lack of favoured orientation when using abstract visual secrets, the vertical and horizontal lines are only identifiable by what is written on them. A quick improvement would then be to switch from "Vote 1/2" or "Vote A/B" to using "Vote 1" on one line and "Vote A" on the other, although it could confuse some voters[19]. A variant useful in certain cases would also be to have "Yes" and "No".

If needed, one could refine the ballot design to make them harder to describe (at the cost of potentially lowering the recognisability). For example, each ballot could have a thicker black border on the edges, to limit the possibility of catching part of someone else's picture if they bend the ballot. The visual secret could also be made circular (with a border covering the rest of the ballot), to make it harder to pinpoint what is in the corners (which was attempted by some subjects). A more complex option is to use a border that is randomly generated in the same style as the general abstract images, but that is common to all the ballots in a pack. This would probably strongly increase the number of ambiguous descriptions, but risks having a similarly strong decreasing effect on the recognisability. Ultimately, a balance needs to be found between recognisability and describability, upon which it is hard to conjecture without further user studies.

[17] To the best of our knowledge, there exists no low-tech verifiable voting system that is fully accessible to the blind.

[18] The impact of candidate order has mostly been observed in large-scale elections where voters are not necessarily familiar with the candidates, and as such could have lower importance in boardroom settings [35].

[19] One small issue is that this is not directly compatible with the texture bumps as the symbols for 1 and A are identical in Braille, unless the complex numerical prefix is added.

Finally, there is the question of the scalability. The subjects in our experiment have shown their ability to remember multiple images. If we were to use 3 independent visual secrets on the same ballot, the probability of fraud detection could potentially reach 99.1% for each ballot (assuming all of the images are unique). An adversary coercing the subject into describing their visual secrets would have at best an 1.5% chance of correctly identifying all the images. This method requires more investigation, as a partial description could be enough to identify the voter depending on how the full set of visual secrets is constructed. Once again, we have balance problems that require more data to be resolved. That said, if a variation on the method could handle sets of 1 000 visual secrets, this kind of vote could happen on a much larger scale, with the verification happening by precinct, each set of visual secret being restricted to its precinct.

7 Concluding remarks

This paper introduced a security primitive called visual secrets, a kind of non-shareable secret that is pure information and does not depend on possessing an item. Its strength comes from two properties :

- the high recognisability of the pictures, with subjects having 80%+ chance of recognising their own secret ;
- the difficulty of unambiguously describing the pictures. No assessor managed to get better than 81% accuracy on the 15-25% of descriptions which they thought were unambiguous.

This primitive allows new possibilities in terms of low-tech protocols that do not require complex sensors. The accuracy figures mean that the visual secrets could be used as is in specific contexts, such as the voting protocol we suggest, or potentially as a replacement for the identifying marks used in other verifiable voting systems inspired by Ron Rivest's ThreeBallot protocol [3, 33]. This would lower the probability of fraud detection from 33% to 28% per ballot, which would be absorbed by the exponential behaviour when detecting fraud on multiple ballots. Moreover, the accuracy could be amplified by the simultaneous use of multiple visual secrets, as discussed previously.

Outside of voting protocols, visual secrets could also be used within authentication mechanisms or online communication protocols. However, this is not trivial as the goals of visual secrets are quite different from picture passwords, their closest equivalent, with describability and short-term memorability being bigger concerns than speed or long-term memorability. As an authentication mechanism, they are then probably suited to very specific use cases. We still hope that this new primitive will inspire the development of systems with improved security and privacy in many different settings.

7.1 Future work and open problems

The data in our experiment has already been released publicly, and there are still a few leads that could be worth investigating (in an exploratory fashion) :

- Our initial attempts at making clustering analyses did not give us any strong insights, but variations in memorability between the images within each series could be compared to a clustering of the descriptions.
- The memorability could be correlated with the time spent on the image description, and inversely correlated with the time spent on the description of other images (as they function as distractor tasks).
- The mobile interface used by 20% of subjects could have had an impact on performance, compared to a full-sized computer screen (potentially not because of the screen but because of how people interacted with the experiment).
- Describability could vary by subject, with some people being better at describing everything unambiguously, or it could mostly depend on the images assigned to the subjects.

Beyond this work on the dataset, this work raises multiple questions about refinements and extensions :

- Could the image recognition process be fooled by images that are very similar, such as iterations on one basis made using generative adversarial networks [9] ?
- What would the performance become if the image series were composed of 100 images or more ?
- We did not measure confidence in the subject's choices when recognising, but it plays a role in the security aspect (in terms of false positives). What would be the recognition performance if we restricted to subjects who are sure of their choice ?
- Would visual secrets be viable with human faces, and how would one correct for demographic variation without knowing the subjects or users in advance ?
- How unambiguous would the descriptions be if we asked the subjects to describe from memory, a few minutes after viewing their pictures (which is closer to the real-life coercion scenario) ?
- How unambiguous would they be if we allowed subjects to view the other pictures ? What if we did so for a limited time, or only for a fraction of the image set ? Could an adversary with some information on the image series create a teachable description method that would increase describability ?

- From a formal standpoint in both classical and quantum complexity, what constraints would allow non-shareable secrets ?
- How sensitive to environmental conditions is the process ? Our study happened *in situ*, but the images were presumably shown and recognised with identical screen settings. Would the performance be affected by the use of printed images or varying luminosity ?
- As relying on sight alone could cause accessibility issues, would it be possible to create tactile secrets (that could be also embossed in a ballot) ? Could an auditory equivalent be viable ? What kind of auditory stimulus would achieve the same recognisability, and would length be a hindrance as sound is a more linear medium?

Finally, our study of VVB is based on the performance of the original visual secret user study. A dedicated usability study on its performance in real life could reveal new intuitions and leads for further improvements.

References

[1] Mathilde Arnaud, Véronique Cortier, and Cyrille Wiedling. Analysis of an electronic boardroom voting system. In James Heather, Steve Schneider, and Vanessa Teague, editors, *E-Voting and Identify*, pages 109–126, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[2] Roman Bednarik, Tomi Kinnunen, Andrei Mihaila, and Pasi Fränti. Eye-movements as a biometric. In Heikki Kalviainen, Jussi Parkkinen, and Arto Kaarna, editors, *Image Analysis*, pages 780–789, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[3] Enka Blanchard and Ted Selker. Origami voting: a non-cryptographic approach to transparent ballot verification. In *5th Workshop on Advances in Secure Electronic Voting*, 2020.

[4] Enka Blanchard, Ted Selker, and Alan T. Sherman. Boardroom voting: Practical verifiable voting with ballot privacy using low-tech cryptography in a single room. preprint at https://hal.archives-ouvertes.fr/hal-02908421/, 2019.

[5] N. Blanchard, Clément Malaingre, and Ted Selker. Improving security and usability of passphrases with guided word choice. In *34th Annual Computer Security Applications Conference, ACSAC*, pages 723–732, 2018.

[6] Hristo Bojinov, Daniel Sanchez, Paul Reber, Dan Boneh, and Patrick Lincoln. Neuroscience meets cryptography: Designing crypto primitives secure against rubber hose attacks. In *21st USENIX Security Symposium*, pages 129–141, Bellevue, WA, 2012. USENIX.

[7] Margaret M. Bradley and Peter J. Lang. Memory, emotion, and pupil diameter: Repetition of natural scenes. *Psychophysiology*, 52(9):1186–1193, 2015.

[8] Bismita Choudhury, Patrick Then, Biju Issac, Valliappan Raman, and Manas Haldar. A survey on biometrics and cancelable biometrics systems. *International Journal of Image and Graphics*, 18, 2018.

[9] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.

[10] Max T. Curran, Jong-kai Yang, Nick Merrill, and John Chuang. Passthoughts authentication with low cost EarEEG. In *IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society – EMBC*, pages 1979–1982. IEEE, 2016.

[11] Valérie Dufour, Michael Coleman, Ruth Campbell, Odile Petit, and Olivier Pascalis. On the species-specificity of face recognition in human adults. *Cahiers de Psychologie Cognitive-Current Psychology of Cognition*, 2004.

[12] Oded Goldreich and Yair Oren. Definitions and properties of zero-knowledge proof systems. *Journal of Cryptology*, 7(1):1–32, 1994.

[13] Jens Groth. Efficient maximal privacy in boardroom voting and anonymous broadcast. In Ari Juels, editor, *Financial Cryptography*, pages 90–104, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[14] Frank Haist, Arthur P Shimamura, and Larry R Squire. On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(4):691, 1992.

[15] Wayne Jensen, Serban Gavrila, Vladimir Korolev, et al. Picture password: A visual login technique for mobile devices. Technical report, National Institute of Standards and Technology, 2003.

[16] Mark H. Johnson. Subcortical face processing. *Nature Reviews Neuroscience*, 6(10):766–774, 2005.

[17] Douglas W. Jones. A brief illustrated history of voting. *University of Iowa Department of Computer Science.*, 2003.

[18] Alexandros Kafkas and Daniela Montaldi. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *The Quarterly Journal of Experimental Psychology*, 64(10):1971–1989, 2011.

[19] J. H. Krantz. Psychological research on the net, 2019.

[20] John H. Krantz and Ulf-Dietrich Reips. The state of web-based research: A survey and call for inclusion in curricula. *Behavior research methods*, 49(5):1621–1629, 2017.

[21] Sheng Li and Alex C. Kot. Attack using reconstructed fingerprint. In *IEEE International Workshop on Information Forensics and Security – WIFS*, pages 1–6. IEEE, 2011.

[22] Feng Lin, Kun Woo Cho, Chen Song, Zhanpeng Jin, and Wenyao Xu. Exploring a secure and truly cancelable brain biometrics for smart headwear. In *Proceedings of the 16th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 1–16, 2018.

[23] Geoffrey R. Loftus. Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 3(4):525–551, 1972.

[24] Patrick McCorry, Siamak F. Shahandashti, and Feng Hao. A smart contract for boardroom voting with maximum voter privacy. In Aggelos Kiayias, editor, *Financial Cryptography and Data Security*, pages 357–375, Cham, 2017. Springer International Publishing.

[25] Shane McCulley and Vassil Roussev. Latent typing biometrics in online collaboration services. In *Proceedings of the 34th Annual Computer Security Applications Conference*, ACSAC '18, pages 66–76, New York, NY, USA, 2018. ACM.

[26] Marnix Naber, Stefan Frässle, Ueli Rutishauser, and Wolfgang Einhäuser. Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of vision*, 13(2):11–11, 2013.

[27] Mizuho Nakajima and Yasushi Yamaguchi. Extended visual cryptography for natural images. In *The 10-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002, WSCG 2002, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, February 4-8, 2002*, pages 303–310, 2002.

[28] Moni Naor and Adi Shamir. Visual cryptography. In Alfredo De Santis, editor, *Advances in Cryptology — EUROCRYPT'94*, pages 1–12, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.

[29] Charles A. Nelson. The development and neural bases of face recognition. *Infant and Child Development: An International Journal of Research and Practice*, 10(1-2):3–18, 2001.

[30] NFB Jernigan Institute. The Braille literacy crisis in America. Technical report, National Federation of the Blind, 2009.

[31] Michael A Nielsen, Isaac L Chuang, and Isaac L Chuang. *Quantum Computation and Quantum Information*. Number 2. Cambridge University Press, 2000.

[32] David Noton and Lawrence Stark. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11(9), 1971.

[33] Ronald L. Rivest and Warren D. Smith. Three voting protocols: Threeballot, vav, and twin. In *Proceedings of the USENIX Workshop on Accurate Electronic Voting Technology*, EVT'07, pages 16–16, Berkeley, CA, USA, 2007. USENIX Association.

[34] Roger N. Shepard. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6:156–163, 02 1967.

[35] Sarah M. Sled. Vertical proximity effects in the California recall election. Technical report, Caltech/MIT Voting Technology Project, 2003.

[36] Ivo Sluganovic, Marc Roeschlin, Kasper B. Rasmussen, and Ivan Martinovic. Using reflexive eye movements for fast challenge-response authentication. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 1056–1067, New York, NY, USA, 2016. ACM.

[37] Tirta Susilo and Bradley Duchaine. Advances in developmental prosopagnosia research. *Current Opinion in Neurobiology*, 23(3):423 – 429, 2013. Social and emotional neuroscience.

[38] Lucina Q. Uddin, Mari S. Davies, Ashley A. Scott, Eran Zaidel, Susan Y. Bookheimer, Marco Iacoboni, and Mirella Dapretto. Neural basis of self and other representation in autism: an fmri study of self-face recognition. *PloS one*, 3(10):e3526, 2008.

[39] Holger Wiese. The role of age and ethnic group in face recognition memory: Erp evidence from a combined own-age and own-race bias study. *Biological Psychology*, 89(1):137 – 147, 2012.

[40] Nicole Wolff, Kathleen Kemter, Stefan R Schweinberger, and Holger Wiese. What drives social in-group biases in face recognition memory? erp evidence from the own-gender bias. *Social Cognitive and Affective Neuroscience*, 9(5):580–590, 2014.