Improving security and usability of passphrases with guided word choice

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

Long passphrases have many uses, such as serving as seeds for high-quality passwords. User-created passphrases are easier to remember, but have been less secure than ones created from words randomly chosen in a dictionary.

This paper attempts to develop a way of making more memorable as well as more secure passphrases. It investigates the effects of creating a passphrase by choosing from a randomly generated set of words presented as an array.

A usability experiment shows that participants using this method are more affected by the word's position in the array than by word familiarity. Passphrases chosen from randomly generated lists achieved 97% to 99% of the maximal entropy in randomly generated passphrases and caused less than half of the memory mistakes.

Prompting a person with random words from a large dictionary is an effective way of helping them make a high-entropy, easy-to-remember passphrase.

ACM Classification Keywords

D.4.6 Security and Protection: Authentication; H.5.2 User Interfaces: User-centered design

Author Keywords

Usable security; Mnemonic phrases; Passwords

INTRODUCTION

Typical uses of passwords have suffered from serious usability and security problems [8, 9]. Low-entropy selection methods, poor memorability, and rules that make passwords difficult to retrieve all reduce their utility. Biometric methods are promising, but they still suffer from many vulnerabilities, typically being hacked within six months of introduction [7, 31, 26, 27]. Biometric security approaches also have an increased risk of unmitigatable leaks about a particular user [30] (as a retina is harder to change than a password). It is time for usable security to become a focus for the usability community.

CHI'16, May 07–12, 2016, San Jose, CA, USA

ISBN 123-4567-24-567/08/06…$15.00

DOI: http://dx.doi.org/10.475/123_4

There is a need for other security alternatives that can handle sharing and changing the access method (such as giving a database password to someone for the evening). Longer lists of words – or passphrases – have been suggested as a possible solution for improving security [29, 32]. Although they were introduced as early as 1982 [25], dozens of years later they still suffer from problems similar to the ones passwords have.

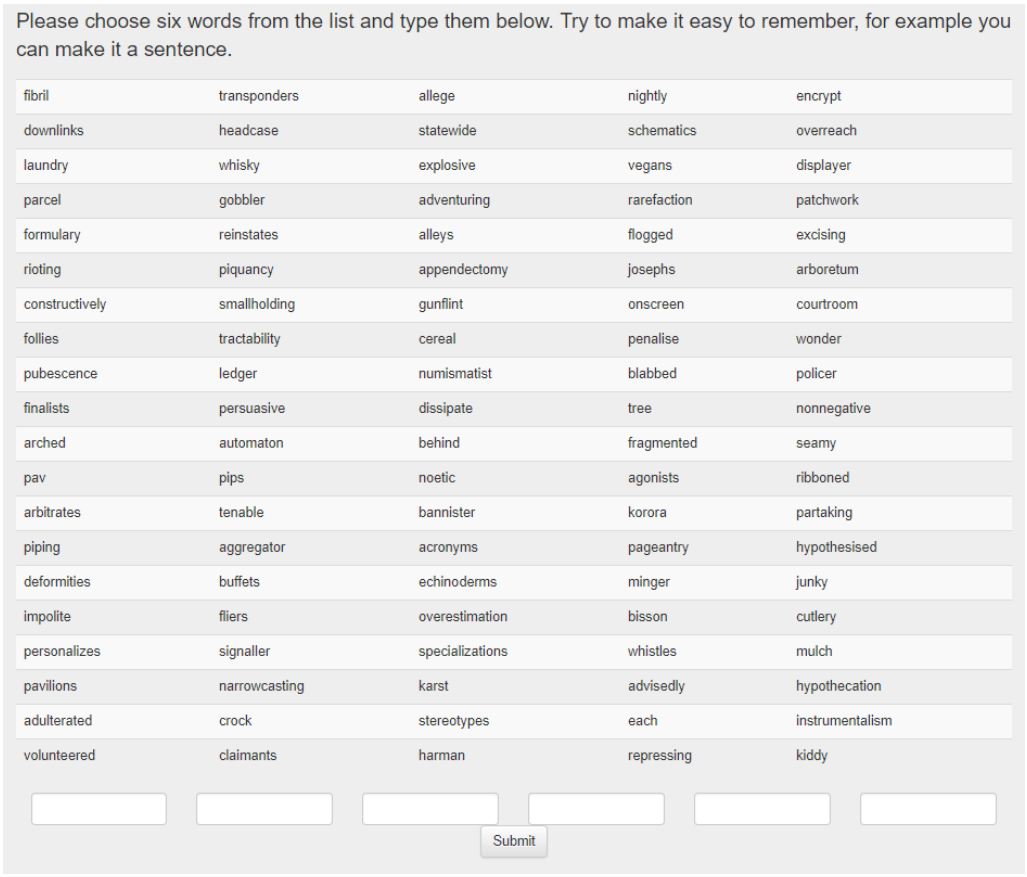
Several studies have shown that passphrases are often made from insecure, linguistically easy-to-crack patterns [18], like song lyrics or famous quotes. A significant number of passphrases created by the Amazon PayPhrase system were easily hackable, with 1.3 percent of accounts being vulnerable to a 20,000-word dictionary of terms used in popular culture [4, 3]. In another study [35], one passphrase method led 2.55% of users to choose the same sentence, and five out of six methods had many occurrences of different users ending up with the same passphrase.

For passphrases to be useful, they need to have high entropy. To achieve this, they have to be not only longer but, more importantly, less predictable. High-entropy passphrases can then be useful for many purposes, such as seeding diverse passwords [2]. As opposed to complex passwords which are hard to remember [16, 34, 24], passphrases benefit directly from our natural abilities to remember sequences of words [1, 21]. They can then serve as a high-entropy source for methods which avoid the pitfalls of password reuse [28, 33, 20].

To avoid users choosing common passphrases, one could take inspiration from standard password practice and draw words uniformly from a dictionary. However, this has two drawbacks: first, the passphrases generated are not individualized and can suffer from low memorability. Second, to make sure that the user knows all the words generated, the dictionary from which the words are drawn must be of limited size.

This paper explores a method of guiding the user to choose their passphrase from an imposed set of random words. A usability experiment explores the factors affecting the word choice, the participants' ability to remember their passphrases, and the type of mistakes they make. Both for native and non-native English speakers, this method leads to highly increased memorability of the passphrases created. Moreover, since the user can choose from several words, the dictionary need not be made of only words that are sure to be known to a user, allowing the use of much bigger dictionaries, leading to an increase not only in usability but also in security.

Figure 1. Screenshot of the word choosing interface



METHOD

An online usability experiment explored the impact of creating a passphrase by choosing words from an array.

Word choice

Presented guide words were drawn uniformly (each word having the same probability) from a dictionary crafted for this purpose. This dictionary is based on the first third of Peter Norvig's 300000 most frequent n-grams [22]. As those 100000 words still included words from other languages such as "unglaublichen" as well as some non-words like "unixcompile", only ones which were also in the SOWPODS (list of admissible words in English Scrabble tournament) were kept. This created a list of the 87691 most frequent English words. Thanks to shared roots in words, there is evidence that most people would know a large majority of them [6, 13]. As the participants chose only 6 words from the array, having a few unknown words in it did not change the process or outcome.

To give participants a real choice, the guide array to choose from was created to be several times the length of the passphrase created from it. It also needed to not be so long that people got confused or took too much time making decisions. For these reasons the size of the word array was chosen to be either 20 or 100, to ensure that the participants could choose from a sufficient number of words they knew, while fitting the whole array on a computer screen.

Protocol

The experiment was hosted on a privately hosted server and accessed remotely over the web. It ran the Scala Play framework and a PostgreSQL database. Participants were shown the following pages, with instructions at the top of each:

- A welcome page that gave participants an overview of the activity and informed them of their rights.
- A question asking their age and another asking the main language they used in everyday life.
- A dynamically generated array of either 20 or 100 words (A/B testing) presented in five columns of either 4 or 20 words. Participants were told to select 6 words, in the order of their choice, and input them in the 6 text-boxes at the bottom of the page. To avoid confusion and separate the guide words from the user-chosen passphrase, the guide array was purposely designed not to line up with the spaces for words in the passphrase below (the interface is shown on Figure 1). Participants were also told to try to make those words memorable, for example, by creating a phrase, rhyme, or sentence that included them.
  A control experiment was run to create a baseline for remembering a 6-word passphrase. Instead of choosing their words from an array, a sequence of 6 randomly generated words was directly given to them, informing them that it had been randomly created.

- A page that repeated the passphrase, then prompted participants to repeat it to themselves until they could remember it.
- A text-box with instructions to type in the first two letters of each of their words. This was meant to help memorize the passphrase and check whether it was memorized.
- A distractor exercise introduced to interfere with their short term memory for passphrases. The idea was to eliminate any short term memory trace that would confound understanding of remembering the passphrases. For this they were shown an array of words that was previously presented to another participant. They were then told to try guessing what words the other participant had chosen.
- A page informing them which if any of their guessed words were correct, and telling them that they could try to guess more passphrases if they wanted, or could continue with the rest of the experiment.
- A page asking them to repeat all six words from their passphrase in the same order, or as many as they could think of if they didn't remember all of them. In case they failed, they were then presented with their original array of words and asked to find all six of them with this clue.
- A page thanking them for their participation and inviting them to encourage others to become participants.

The experiment collected of the following data for analysis:

- Any information entered in the text-boxes.
- All the words and arrays shown to the participants.
- Time spent on each page.
- List of (keystroke/timestamp) couples.

To make sure that no one would try the experiment multiple times to improve their performance, IP addresses associated with the participants were temporarily kept. A single occurrence of a second try by a participant was detected and was excluded from the database.

Participant selection

All the participants were volunteers, and were informed of the length of experiment and that they could quit at any point. They were told that it was an opportunity to help them test their memory and for us to understand how people typed, and that their typing would be monitored. For privacy, minimal demographic data was collected, corresponding to aspects that would be relevant to analysing results.

Recruitment of volunteers

Initial volunteers were recruited through John Krantz's Psychological Research on the Net website[17] which promotes and indexes experiments of this kind. These volunteers were encouraged to invite others to participate. Due to this, a significant proportion of later volunteers were probably recruited through social networks.

RESULTS

The results focus on how participants chose words and how different variables affected their choices, on their ability to remember the words chosen, and on how they guessed other people's words.

Word selection

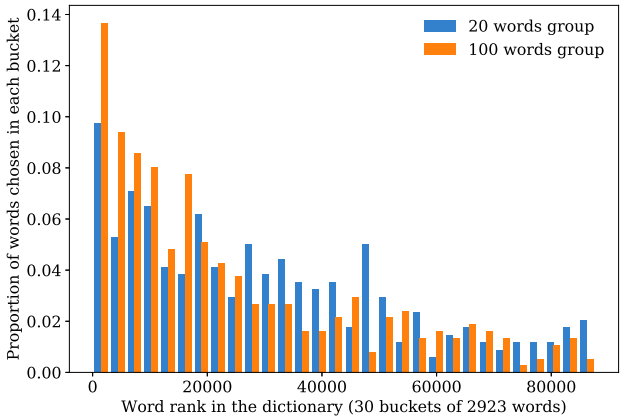
Based on problems recognized in other passphrase studies, an original hypothesis was that word choice would be influenced by three behaviours:

- *Semantic*: participants might choose words that are more frequently used (and with which they were more familiar);
- *Syntactic*: participants might choose words that are compatible with others they chose, to create a sentence with a common structure.
- *Positional*: participants might choose words that are either among the first they read (on the top left corner), or closest to the input fields;

Results below showed that the first two are much weaker than could be expected, with most of the choosing bias caused by positional effects.

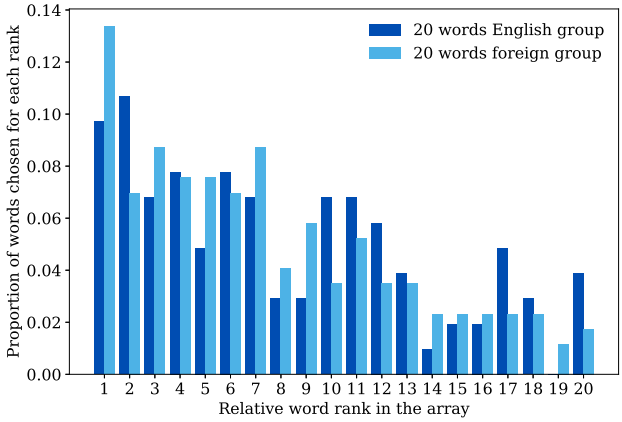
Semantic effects

Figure 2. Frequencies of the words chosen by each group as a function of their rank in the dictionary, by buckets of 2923 words



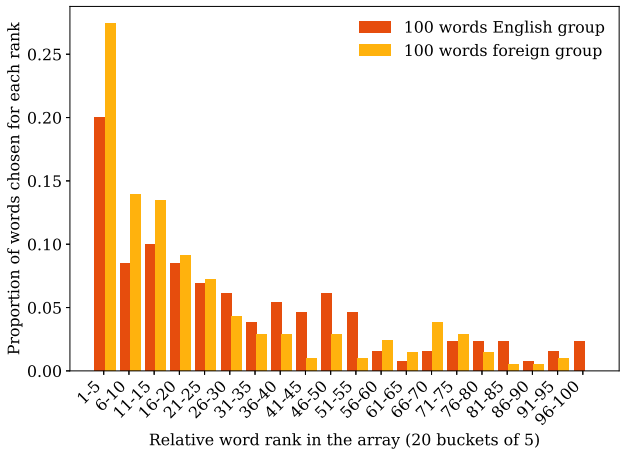
The histograms in Figure 2 show the distribution of the words chosen depending on their position in the dictionary (which is sorted by decreasing frequency), for each of the two main groups.

Figure 3. Relative word frequencies for the 20 words group



Figures 3 and 4 show the distribution of words chosen depending on their frequency relative to the frequencies of the words shown to the participant. F(i) is equal to the number of times the i-th most frequent word in the array was chosen.

Figure 4. Relative word frequencies for the 100 words group

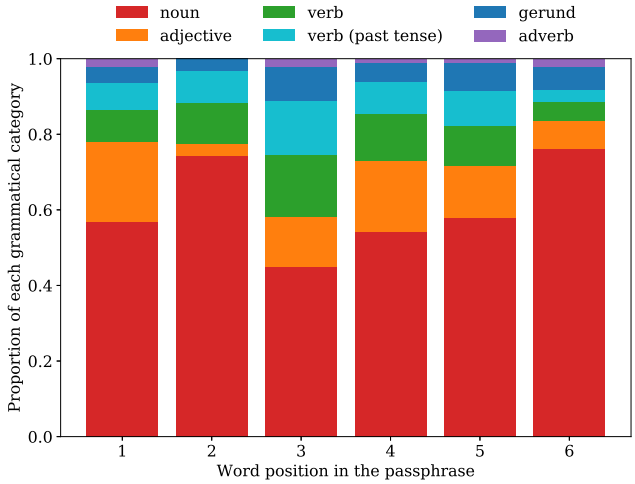


Syntactic effects

While using common sentence structures might help memorization, most participants did not seem to follow such a strategy in the passphrases they created. Even when limiting the analysis to the four broadest grammatical categories (noun/verb/adjective/adverb), most sentence structures were unique, with 65 structures seen only once out of 99 passphrases. 10 structures were seen twice and 3 were seen thrice. The only relatively common structure, which was present 8 times, corresponds to a sequence of 6 nouns.

Figure 5 shows how present each grammatical category was in each position in the passphrase. For example, adjectives are less present as a second word than as a first. There is some imprecision as words can fit in multiple categories (e.g. *scars* as a noun or a verb). Nouns seem overrepresented, but this is consistent with their frequency in the dictionary ($\approx 60\%$).

Figure 5. Repartition of grammatical categories by position in passphrase



In the 20 words group, 12 people created passphrases that made some amount of semantic sense and followed English syntax. 13 passphrases could make some sense but had unusual or incorrect syntax. 22 appeared to be six randomly ordered words. In the 100 words group, only 6 people created passphrases that made some amount of sense and were syntactically correct. 15 made passphrases that could make some sense, and 30 had passphrases that seemed randomly ordered.

Positional effects

The heat maps in Figure 6 and 7 show how the position of a word in the array shown greatly affected the probability that it would get chosen. The numbers correspond to the percentage of participants who chose the word in that cell, with a deeper red indicating a higher percentage.The numbers under and beside the heat maps represent the total number of words chosen by line and column.

Figure 6. Heatmap indicating the percentage of participants choosing the word in each cell for the 100 words group

32.6	17.9	29.3	31.0	19.6	80
11.4	13.0	4.9	9.8	16.3	34
6.5	17.9	14.7	17.9	14.7	44
4.9	8.2	9.8	11.4	13.0	29
6.5	11.4	8.2	9.8	6.5	26
6.5	6.5	1.6	9.8	8.2	20
6.5	8.2	1.6	6.5	1.6	15
6.5	0.0	6.5	11.4	1.6	16
3.3	0.0	3.3	4.9	8.2	12
4.9	3.3	9.8	1.6	6.5	16
8.2	1.6	1.6	6.5	1.6	12
0.0	3.3	4.9	1.6	4.9	9
1.6	6.5	0.0	1.6	1.6	7
3.3	1.6	3.3	6.5	3.3	11
6.5	1.6	4.9	3.3	0.0	10
1.6	0.0	1.6	1.6	4.9	6
1.6	3.3	1.6	0.0	3.3	6
1.6	1.6	1.6	0.0	1.6	4
1.6	1.6	1.6	3.3	0.0	5
1.6	1.6	1.6	1.6	3.3	6
72	67	69	86	74	368

Figure 7. Heatmap indicating the percentage of participants choosing the word in each cell for the 20 words group

72.6	53.5	47.8	47.8	40.1	137
47.8	45.9	24.8	32.5	22.9	91
32.5	24.8	19.1	11.5	11.5	52
13.4	17.2	13.4	7.6	13.4	34
87	74	55	52	46	314

Memorization

After choosing a passphrase and performing a distractor task, users were asked to recall their original passphrase; 48% of

participants managed to get all 6 words in their passphrase without any errors. An additional 32% remembered their words but made a typo somewhere. The table in Figure 8 sums up the recall success rate and types of errors when recalling passphrases. A full explanation is given below. The rates are calculated separately for the 20 words group, the 100 words group, and the control group. They are also split between the first and the second section, in which (except for the control group) participants were reminded of their original array of words.

Figure 8. Total number of errors by type

Section	Correct	Typo	Variant	Order	Miss	Wrong
1:20	19/47	6	8	6	26	5
1:100	26/51	10	5	3	16	4
Control	6/26	11	11	10	31	12
2:20	14/29	1	2	8	0	3
2:100	15/26	4	2	3	1	4

The following mistakes and errors showed up in participant recall of newly chosen passphrases:

- Typos are simple one-letter errors or exchanges between two adjacent letters.
- Variants are like typos in that they are related English words. Most of those were verbs where the participant added or removed an 's' (or less frequently an 'ed' or 'ing').
- Orders are errors where at least two words are exchanged in the passphrase.
- Misses are words that are entirely missing from the passphrase entered.
- Wrong words are ones that have no relation to any word in the original passphrase.

Overall, only 6% of the 51 people that were guided with 100 words had forgotten two or more passphrase words. 19% the 47 of people who were shown 20 words had forgotten two of them or more (that number reaches 62% in the control group). When comparing the number of people who correctly remembered their whole passphrase, the 100 words group is superior to the control group ($p < 0.02$). This effect is magnified when comparing not the participants but the words directly. The words in passphrases made from the 100 words array were better remembered than those made from the 20 words array ($p < 0.03$). Similarly, the words in passphrases made from the 20 words array were themselves much better remembered than the ones given to participants in the control group ($p < 10^{-4}$).

The creation of sentence-like passphrases had no statistically significant impact on overall success rate in either group, with a small decrease in misses and a increase in false words ($p > 0.05$).

We can also restrict the analysis to those who remembered the passphrase correctly just after making it in the first exercise (the one asking them to type the first two letters of each word). This is shown in Table 9.

Figure 9. Errors by type for participants with correct first exercise

Section	Correct	Typo	Variant	Order	Misses	Wrong
1:20	19/41	4	8	2	14	5
1:100	26/45	9	5	1	14	1
Control	6/15	5	4	1	7	2

Remarks

The preceding error tables do not take into account four anomalous behaviours. Two participants (one in each group) made a typo in their original passphrase (phrases were only counted as correct when typed without the typo). One participant, when asked for the first two letters for each word in their passphrase, typed random letters on their keyboard, and one typed something that looked like the requested twelve-character string with lots of mistakes. Both of those were in the 20 words group and were not counted in the analysis. One participant in the 100 words group also double-clicked on the next button and was taken directly to the second try and shown their array of words. Finally, four participants in the control group showed no attempt to recall their passphrase, responding with random words (and in one case not even filling the 6 phrase positions), and were removed from the dataset. Including these would have only strengthened the results that guided word choice helps.

Language

People that identified their primary language as English were balanced between the two groups (25 of 51 in the 100 words group and 26 of 47 in the other). Language did not show a statistically significant effect: 21 out of 51 people who indicated English as their primary language were correct on the first try, as were 23 out of the 48 who indicated another language. Native English speakers had more misses (29 against 13) and an equal share of wrong words.

Time

No statistically significant advantage was shown for participants who spent more time designing their passphrase. People who recalled their words perfectly appeared to take 5-10% longer on average, but 10-15% less time for the median, showing no clear effect.

Some of the participants disabled the JavaScript functions needed to record the time taken[1].

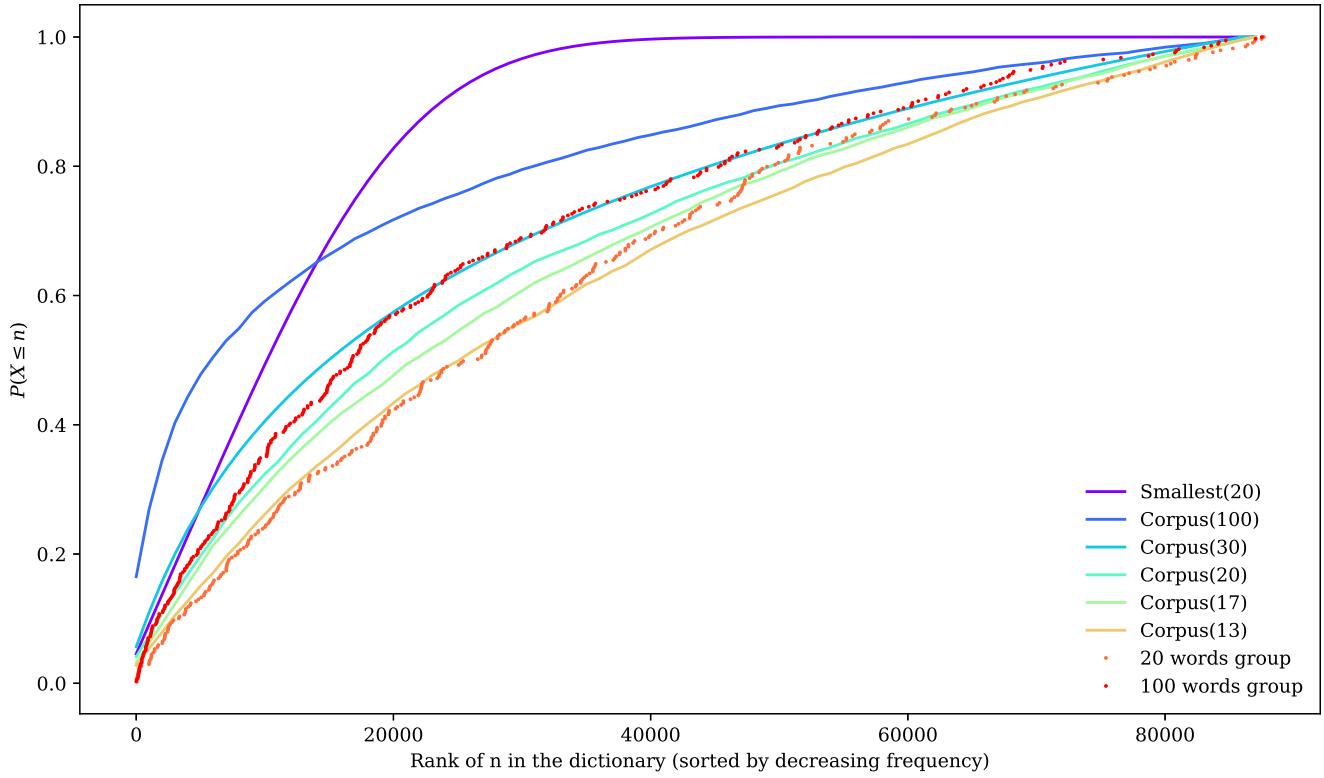
Guessing

Most participants only tried to guess a single other passphrase, but 21 participants tried to guess between 2 and 4 paraphrases. Words that were in the original passphrase with a minor modification (such as a typo) were counted as correct in this exercise, like those that were only in the wrong place.

On average, participants could guess 0.85 out of 6 words chosen by another person from 100 word arrays, and 2.15 out of 6 words from 20 word arrays. this is significantly higher than a random guess (which would be close to 0.36 and 1.80), but lower than an educated guess that would focus on the words appearing in the upper left corner of the array.

[1] These are disabled by default on most Apple iPhones.

Figure 10. Cumulative distribution function of indices of words chosen for both groups and 6 models



Demographic information

Age

The participants' ages showed a large variation, going from 16 to 69, with a notable concentration around 24. The average was 31 years old, and the median 25.

Language

51 participants wrote down English as their main language. The next two groups were French (28) and Hebrew (14), followed by Arabic, Norwegian, Russian, and Romanian.

Dropout rates

148 people went through the participant information page. The experiment was actually completed by 100 people (including the participant who did it twice). People who were shown the array of 20 words immediately quit more frequently than the ones who were shown the 100 words array (58 out of 76 continued to the next page, against 64 out of 72, $p < 0.05$).

In the control group, out of 56 initial participants who got to the welcome page, only 30 finished the experiments, with 19 stopping right after getting their passphrase.

STATISTICAL MODELLING

The effect of participant choice on the entropy of passphrases was tested. Models simulated in Python were used to analyse the word choice of participants when presented with arrays of words, with three main strategies:

The *Smallest*(n) strategy corresponded to picking the six most frequent words presented in the array of n words.

The *Uniform*(n) strategy being equivalent to sampling random words uniformly from a dictionary of size n, entropy was computed exactly. The table in Figure 11 shows the entropy per word for each of the described strategies, as well as a few others. The *87691* here corresponds to our described dictionary of *87691* words, and *300000* to Norvig's more complete dictionary of *300000* words.

Finally, the *Corpus*(n) strategy corresponded to picking each word from an array of n words according to a distribution where the probability to pick each word w is a function of $f(w)$, its frequency in the language. This corresponds to models for how the distribution of words in the English language – among others – is biased in many different corpora of texts. The model used here is Zipf's law[12], stating that the probability of choosing a word $p(w)$, is inversely proportional to its rank in the frequency list:

$$p(w) \propto \frac{1}{rank(w)}$$

Informally, the 100^{th} most used word is chosen with a frequency about twice the frequency of the 200^{th} most used word.

10^9 simulations were run for both 20-word and 100-word arrays with each strategy to estimate the probabilities – and the entropy. For entropy E, the formula to compute it is:

$$E = -\sum_i p_i ln(p_i)$$

In this formula, p_i is the position in an ordering of word frequency that the word occupies in the array of words.

Figure 11. Strategies and entropy

Strategy	Entropy (bits)
Uniform(87691)	16.42
Corpus(13)	16.25
Corpus(17)	16.15
Corpus(20)	16.10
Corpus(30)	15.92
Corpus(100)	15.32
Uniform(10000)	13.29
Smallest(20)	12.55
Uniform(5000)	12.29
Uniform(2000)	10.97
Smallest(100)	10.69
Corpus(300000)	8.94
Corpus(87691)	8.20

A much bigger sample would be needed to exactly compute the entropies of user behaviours. However, experimental entropy can be bounded by using distributions for known entropies. As such, the cumulative distribution functions for the experimental groups and models were computed. Although they do not make direct strategic sense, we included *Corpus*(17) and *Corpus*(13) in Figure 10, as they bound the observed curve for the 20 words group.

Experimental values for the 20 words group are slightly above *Corpus*(17) around the 50000$^{\text{th}}$ word. An upper bound of *Corpus*(20) could be chosen, but there is a strong argument for using *Corpus*(17). This is more affected by the values of the high p_i, as the function changes less as it gets to the least common words (making it concave). As the p_is shown in Figure 10 are also sorted in decreasing order, it means a small bump in word choice in the first part of the curve is more than compensated by the lack of a bump in the second part. As such, it is reasonable to infer that the entropy corresponding to participants' behaviours in this group is between *Corpus*(13) and *Corpus*(17).

A slightly tighter fit can be obtained by taking not the simplest Zipf's formula but the more general one, with, for $\beta > 1$:

$$p(w) \propto \frac{1}{(rank(w))^{\beta}}$$

In such a case, setting $\beta = 1.35$ makes *Corpus-Zipf*(13) a tighter fit than previous curves, giving an entropy of 16.19 bits[2]. However, the presence of noise in the data means that a search for a more accurate model would be premature.

[2]One could also use a Zipf-Mandelbrot model [23], but the additional parameter would be hard to validate accurately without having a sample of at least 10000 participants

DISCUSSION

The above results demonstrate that creating a passphrase by choosing words from an array of random words is much more memorable than being given a passphrase. While past studies have shown that choosing familiar words for passphrases led to huge entropy reductions, our technique obviated this. The entropy cost due to the choice in our system is negligible, staying between 1% and 3% depending on array size. The method also allows the use of a larger dictionary to choose known words from, leading to much higher entropy per word in the end.

Multiple surprising behaviours were observed, confirming certain hypotheses and refuting others. Firstly, the participants' choices were greatly influenced by the positions of the words in the arrays presented to them. In the 20 words group, this led 73% of them to choose the word in the upper left corner, instead of the expected 30%. Variations of one order of magnitude between different positions in the array were found in both groups. Although there was a very strong bias for the top lines, the left-wise bias was only observed in the 20 words group[3].

This position effect is greater than the tendency to choose familiar words, with a linguistic bias even weaker than the one in the English language (as predicted by Zipf's law). We observed that *Corpus*(13-17) might be better fits than *Corpus*(20) for the 20 words group. This can be explained by the fact that its participants nearly ignored the bottom right corner of the array. Similarly, only 20.6% of the words chosen by the other group were in the bottom half of the array, and their choices are much more consistent with *Corpus*(30) than with *Corpus*(100).

One might expect participants to use mnemonics and create sentence-like passphrases that used common patterns, but the only syntactic pattern that appeared more than thrice was 8% of participants choosing three nouns in a row. As nouns form the bigger share of the dictionary used, even those passphrases are secure, reducing the entropy by at most 5 bits (out of more than 96). Moreover, having more words to choose from did not increase the tendency to create syntactically correct sentences but reduced it instead.

Surprisingly, word choice patterns held across the range of proficiencies in English. Those results are true not just for native English speakers but also for people for whom it is a second language, with barely noticeable differences in word choice and no significant difference in memorability.

The task where participants were supposed to guess each other's words had a purpose beyond distracting and impairing their memory. We were hoping to find whether a simple strategy could explain the participants' choices, in which case some would get very good results. The absence of such successful participants shows that if a general strategy to explain participants' choices exists, it has eluded both us and them.

[3]Unfortunately we did not compare multiple ways of presenting a word field. While a linear presentation of the choice words was not tested, it would be likely be much less user-friendly.

CONCLUSION

This paper shows that guiding people to choose from an array of random words can build a high-entropy, highly memorable passphrase. Choosing from an array successfully stops people from choosing easy-to-guess favorite words. The reduction in entropy when compared with random passphrase generation is offset by the improved usability and memorability that come with choice. Since people can choose words they know, larger dictionaries can be used. Final entropy by word then exceeds the levels of random passphrase generation by $20-30\%$.

The main predictor of word choice from the random word array was not how familiar a word was to the participant, but its position in the array. This was true even for non-native speakers. Globally, letting people choose from an array gave between 97% and 99% of the maximal entropy achievable, depending on the size of the array.

Bigger array size (giving more choice to the participant) was linked to improved memorability of the passphrase. With a 100-word array, 94% remembered at least 5 of their 6 words the first time they were asked to recall, despite performing a distractor task. With actual use, remembering the passphrase should be easy.

The advantage of selecting from an array of random words is that it gives users an easy way to create secure and memorable passphrases, with only a random generator and a dictionary. Security of the guiding array is achieved by only generating it when needed, locally on the user's machine. This secret array to select from can be produced on any machine, as the dictionary itself is smaller than 300KB.

This work demonstrates that large improvements can be achieved in passphrase usability while increasing their entropy. Below are a few suggestions for usability/security questions left to test for the guided passphrase scenario:

- Can putting high-frequency words farther from the top of the array compensate for the positional advantage?
- Can other visual presentations, such as word clouds, make word choice from the presented set even more uniform?
- Would the high uniformity of word frequency be as or even more successful with even larger dictionaries, for example with the SOWPODS and its 276663 words?
- Is choosing from 100 words more memorable than choosing from 20 because of more personalized choices, more familiar words, or is it due to another reason? Does memorability continue to increase with arrays of more than 100 words? What is the nature of the trade-off, and must there be a compromise between entropy and memorability?
- How does the size of the array affect reading patterns and word choice, and is a difference in reading pattern the source of the left-wise bias in the 20 words group?
- Are there even better ways to make extremely memorable high-entropy passphrases?

We expect and hope that making easy-to-remember passphrases can improve security for many purposes. With these demonstrations, we hope to inspire more work that will make secure passphrases that are as easy to remember as the song you can't get out of your head.

REFERENCES

1. Alan Baddeley and Graham James Hitch. 1974. *Working memory*. Vol. 8. Academic Press, 47–90.
2. Jeremiah Blocki, Manuel Blum, Anupam Datta, and Santosh Vempala. 2014. Towards Human Computable Passwords. *arXiv preprint arXiv:1404.0024* (2014).
3. J. Bonneau. 2012. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *2012 IEEE Symposium on Security and Privacy*. 538–552. DOI:http://dx.doi.org/10.1109/SP.2012.49
4. Joseph Bonneau and Ekaterina Shutova. 2012. Linguistic properties of multi-word passphrases. In *International Conference on Financial Cryptography and Data Security*. Springer, 1–12.
5. Sacha Brostoff and M. Angela Sasse. 2000. *Are Passfaces More Usable Than Passwords? A Field Trial Investigation*. Springer London, London, 405–424. DOI: http://dx.doi.org/10.1007/978-1-4471-0515-2_27
6. Marc Brysbaert, Michaël Stevens, Paweł Mandera, and Emmanuel Keuleers. 2016. How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Frontiers in Psychology* 7 (2016), 1116. DOI:http://dx.doi.org/10.3389/fpsyg.2016.01116
7. Kai Cao and Anil K. Jain. 2016. Hacking Mobile Phones Using 2D Printed Fingerprints.
8. Lorrie Faith Cranor. 2014. What's wrong with your pa$$word. https://www.ted.com/talks/lorrie_faith_cranor_what_s_wrong_with_your_pa_w0rd. (2014).
9. Lorrie Faith Cranor. 2016. Time to rethink mandatory password changes. (2016). https://www.ftc.gov/news-events/blogs/techftc/2016/03/time-rethink-mandatory-password-changes
10. Leilei Gao and Itamar Simonson. 2016. The positive effect of assortment size on purchase likelihood: The moderating influence of decision order. *Journal of Consumer Psychology* 26, 4 (2016), 542–549.
11. S. Garfinkel and H.R. Lipford. 2014. *Usable Security: History, Themes, and Challenges*. Morgan & Claypool Publishers. https://books.google.fr/books?id=HPS9BAAAQBAJ
12. Le Quan Ha, E. I. Sicilia-Garcia, Ji Ming, and F. J. Smith. 2002. Extension of Zipf's Law to Words and Phrases. In *Proceedings of the 19th International Conference on Computational Linguistics - Volume 1 (COLING '02)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1–6. DOI: http://dx.doi.org/10.3115/1072228.1072345
13. George W Hartmann. 1946. Further evidence on the unexpected large size of recognition vocabularies among college students. *Journal of educational psychology* 37, 7 (1946), 436.

14. Yasser M. Hausawi and William H. Allen. 2014. *An Assessment Framework for Usable-Security Based on Decision Science*. Springer International Publishing, Cham, 33–44. DOI: http://dx.doi.org/10.1007/978-3-319-07620-1_4

15. Blake Ives, Kenneth R. Walsh, and Helmut Schneider. 2004. The Domino Effect of Password Reuse. *Commun. ACM* 47, 4 (April 2004), 75–78. DOI: http://dx.doi.org/10.1145/975817.975820

16. Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L. Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. 2011. Of Passwords and People: Measuring the Effect of Password-composition Policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2595–2604. DOI: http://dx.doi.org/10.1145/1978942.1979321

17. JH Krantz. 1998. Psychological research on the net. *WWW document* (1998).

18. Cynthia Kuo, Sasha Romanosky, and Lorrie Faith Cranor. 2006. Human selection of mnemonic phrase-based passwords. In *Proceedings of the second symposium on Usable privacy and security*. ACM, 67–78.

19. Micah Lee. 2015. Passphrases that you can memorize – but that even the NSA can't guess. https://theintercept.com/2015/03/26/passphrases-can-memorize-attackers-cant-guess. (2015).

20. Peter Lipa. 2016. The Security Risks of Using "Forgot My Password" to Manage Passwords. https://www.stickypassword.com/blog/the-security-risks-of-using-forgot-my-password-to-manage-passwords. (2016). Accessed: 2017-12-18.

21. George A Miller. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81.

22. Peter Norvig. 2009. Natural language corpus data. *Beautiful Data* (2009), 219–242.

23. Richard Charles Oldfield. 1968. *Language: selected readings*. Vol. 10. Penguin.

24. Denise Ranghetti Pilar, Antonio Jaeger, Carlos F. A. Gomes, and Lilian Milnitsky Stein. 2012. Passwords Usage and Human Memory Limitations: A Survey across Age and Educational Background. *PLoS One* 7, 12 (05 Dec 2012), e51067. DOI: http://dx.doi.org/10.1371/journal.pone.0051067 PONE-D-12-21406[PII].

25. Sigmund N Porter. 1982. A password extension for improved human factors. *Computers & Security* 1, 1 (1982), 54–56.

26. P. Venkata Reddy, Ajay Kumar, S. Rahman, and Tanvir Singh Mundra. 2008. A New Antispoofing Approach for Biometric Devices. *IEEE transactions on biomedical circuits and systems* 2 4 (2008), 328–37.

27. Virginia Ruiz-Albacete, Pedro Tome-Gonzalez, Fernando Alonso-Fernandez, Javier Galbally, Julian Fierrez, and Javier Ortega-Garcia. 2008. *Direct Attacks Using Fake Images in Iris Verification*. Springer Berlin Heidelberg, Berlin, Heidelberg, 181–190. DOI: http://dx.doi.org/10.1007/978-3-540-89991-4_19

28. Sean M. Segreti, William Melicher, Saranga Komanduri, Darya Melicher, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Michelle L. Mazurek. 2017. Diversify to Survive: Making Passwords Stronger with Adaptive Policies. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 1–12. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/segreti

29. Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2014. Can Long Passwords Be Secure and Usable?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2927–2936. DOI: http://dx.doi.org/10.1145/2556288.2557377

30. Koen Simoens, Pim Tuyls, and Bart Preneel. 2009. Privacy weaknesses in biometric sketches. In *Security and Privacy, 2009 30th IEEE Symposium on*. IEEE, 188–203.

31. Daniel F. Smith, Arnold Wiliem, and Brian C. Lovell. 2015. Face Recognition on Consumer Devices: Reflections on Replay Attacks. *IEEE Transactions on Information Forensics and Security* 10 (2015), 736–745.

32. Umut Topkara, Mikhail J. Atallah, and Mercan Topkara. 2007. Passwords Decay, Words Endure: Secure and Re-usable Multiple Password Mnemonics. In *Proceedings of the 2007 ACM Symposium on Applied Computing (SAC '07)*. ACM, New York, NY, USA, 292–299. DOI:http://dx.doi.org/10.1145/1244002.1244072

33. Rick Wash, Emilee Rader, Ruthie Berman, and Zac Wellmer. 2016. Understanding Password Choices: How Frequently Entered Passwords Are Re-used across Websites. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 175–188. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/wash

34. Jeff Yan, Alan Blackwell, Ross Anderson, and Alasdair Grant. 2004. Password Memorability and Security: Empirical Results. *IEEE Security and Privacy* 2, 5 (Sept. 2004), 25–31. DOI:http://dx.doi.org/10.1109/MSP.2004.81

35. Weining Yang, Ninghui Li, Omar Chowdhury, Aiping Xiong, and Robert W. Proctor. 2016. An Empirical Study of Mnemonic Sentence-based Password Generation Strategies. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*. ACM, New York, NY, USA, 1216–1229. DOI: http://dx.doi.org/10.1145/2976749.2978346