IDC School of Design
अभिकल्प विद्यालय

IIT Bombay

Balancing Graphical Pin-Based Passwords through Locally-Tailored Celebrity Recognition in Kanpur

Submitted in partial fulfillment of the requirements of the degree of Masters of Design

By:

Sparsh Gupta

22M2252, M.Des Interaction Design 2022-24

Supervisor:

Prof. Anirudha Joshi

IDC School of Design

Indian Institute of Technology, Bombay

2024

Approval Sheet

Interaction Design Project 2 titled “Balancing Graphical Pin-Based Passwords through Locally-Tailored Celebrity Recognition in Kanpur” by Sparsh Gupta (Roll number 22M2252) is approved for partial fulfillments of the requirements for the degree of “Masters in Design” in Interaction Design at Industrial Design Center, Indian Institute of Technology, Bombay.

Guide: Digital Signature
Anirudha N Joshi (i98081)
12-Aug-24 05:51:31 PM

Chairperson: 

Internal Examiner: 

External Examiner: 

Declaration

I declare that the written document represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misinterpreted or fabricated or falsified any idea/fact/ source in my submission. I understand that any violation of the above will be cause for disciplinary action by the institute and can also evoke penal action from the sources which thus not properly cited or from whom proper permission has not been taken when needed.

Sparsh Gupta

22M2252

Industrial Design Centre

IIT Bombay

Acknowledgement

I extend my deepest gratitude to Prof. Anirudha Joshi for his exceptional guidance and unwavering support throughout the duration of this project. His expertise and insights have been invaluable in shaping the course of my work. I would also like to express my sincere appreciation to the professors of IDC, whose teachings have greatly contributed to my academic and professional growth.

A special thank you goes to my family – my mother and father, for their endless love and encouragement, and my brother Ishaan, who not only sacrificed convenient transportation to college, as I took our only vehicle for data collection, but also assisted me diligently in data collection. His enthusiasm for learning and his support in my project endeavors have been a constant source of motivation.

My heartfelt thanks to my friends Yash, Tarun, and Prince, who have been pillars of support throughout this journey. Their constructive criticism and the mini juries, where I had the opportunity to explain my project in detail, were instrumental in refining my work and broadening my perspective.
This project was not just a pursuit of academic excellence but a journey enriched by the support and guidance of remarkable individuals. Thank you all for being an integral part of this journey.

Abstract

This research undertakes a novel approach to enhancing the security of graphical pin-based passwords through a systematic exploration of celebrity recognition within the locale of Kanpur, Uttar Pradesh. Graphical passwords, leveraging easily recognizable celebrity faces, aim to mitigate common issues related to textual passwords such as weak password selection and vulnerability to attacks. However, an inherent imbalance exists due to varied recognition levels of celebrities, necessitating a methodologically robust system to ensure equitable representation and selection probability.
A structured, recognition-based interview protocol will be deployed to ascertain a diverse set of recognizable celebrities, traversing domains like sports, entertainment, and music, amongst others, and including both local and international personalities. A series of open-ended and domain-specific questions will be utilized, with the order and domains being randomized to control for order effects and priming, thus ensuring data richness and reliability.
Data will be subsequently processed to assign popularity scores to each celebrity, based on their frequency of mention and the order in which they were mentioned. The distribution of these scores was analyzed to identify and eliminate extremes, thus formulating balanced grids.

Criteria for grid balance were established by leveraging statistical measures such as mean, standard deviation, and entropy, providing a quantifiable methodology to assess the balance of a grid.

This locally-tailored approach to creating balanced celebrity grids in graphical pin-based passwords offers enhanced security and user convenience. By considering local preferences and recognition levels, this project provides insights and methodologies that can be extrapolated to diverse locales, contributing significantly to the development of more secure and user-friendly authentication systems.

Content

Introduction ——— ১

Taxonomy of Passwords ১

Graphical Passwords ২

Vulnerabilities ৪

Types of Attacks ৬

Graphical Pin-based Passwords ৫

Grid Balancing ৮

Shoulder Surfing ১০

Study Details ——— ১১

Ideation ——— ১২

Ideas ১২

Issues ১৬

Initial data collection approach ১৮

Plans for evaluation ——— ১৯

1st Pilot Deployment ——— ২১

Learnings ২১

Recognition vs Recall ২২

Curated Protocol ২৩

2nd Pilot Deployment ——— ২৬

Learnings ২৬

Finalised Method ——— ২৫

Continuous Adaptability ——— ২৮

Final Deployment Plan ——— ২৯

Final Deployment ——— ৩০

Phase 1 ৩০

Continuously adapting corpus ৩১

Phase 2 ३२
Eliminating Duplicates ३३
Calculation of R_i ३३
Ranking celebrities ३४

Creating Grids ३५

Distribution of R_i ३५
Trimming the list ३५
Creation of balanced grids ३६
Balanced Grids created using this method ३७

Evaluation of grids ३९

Testing App ३९
Experiment Design ४०

Results after data collection ४१

Selection Heatmaps ४२

Comparing Standard Deviations ४५

Calculating Sharron's Entropy ४६

Limitations and Future Scope ४९

Conclusion ५०

Miscellaneous ५१

Positioning of the project ५१
Motivations ५१
Study suggestions for shoulder surfing ५२
References ५५

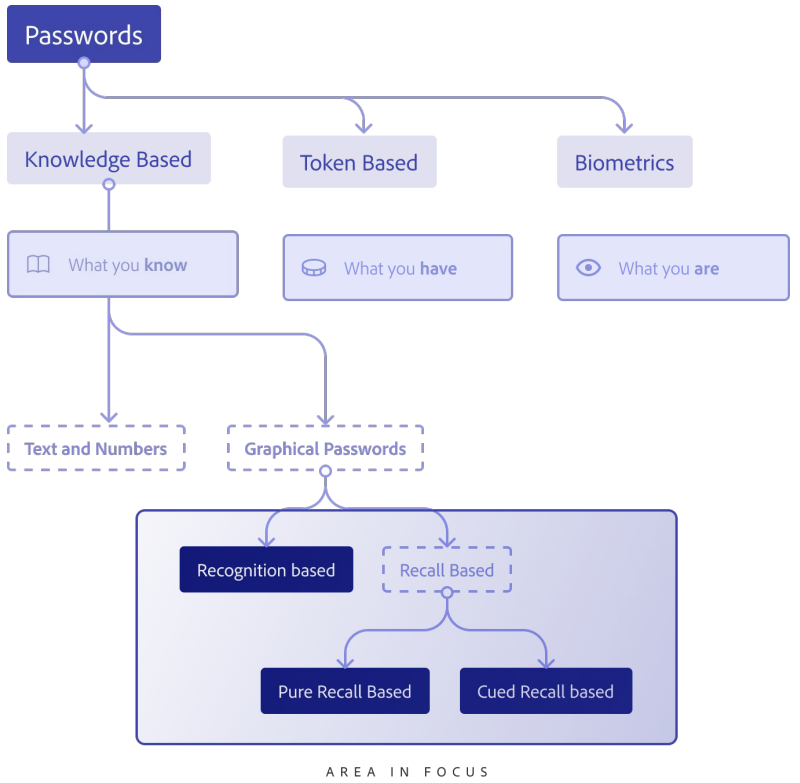
Introduction

Authentication is a crucial component in computer security, especially as our reliance on digital infrastructure continues to grow, necessitating heightened security measures. Various authentication methods have been explored in academic literature, each presenting its unique advantages and disadvantages. Passwords can be classified in 3 categories.

- Knowledge based (What you know)
- Token based (What you have/ possess)
- Biometric based (What you are)

Text-based passwords, which are a knowledge based mode of authentication, remain the most prevalent due to their simplicity of input. However, they come with their own set of challenges.

Ideally, Nonetheless, text-based passwords that are easy to remember are often short or use common words, making them susceptible to dictionary attacks. Passwords containing personal details are also easy to remember but can be easily



Taxonomy of passwords

guessed by acquaintances or attackers who have access to the user's information. The issue stems from the difficulty in remembering passwords, leading to other related issues like password reuse (across platforms), sharing (across users), and the selection of weak passwords. Then they are also more susceptible to other forms of attacks like key-logging.

Graphical Passwords

Graphical passwords have been emerging as an alternative for text-based passwords, capitalizing on the human propensity to remember images better than text. This authentication method necessitates users to engage with visual components to secure access, which might include choosing images in a particular sequence or creating patterns. The fundamental benefit of graphical passwords is that individuals typically find visual data more retainable and easier to remember compared to sequences of letters and numbers, minimizing the chances of forgetting passwords.

Graphical passwords can be further classified into 3 categories:

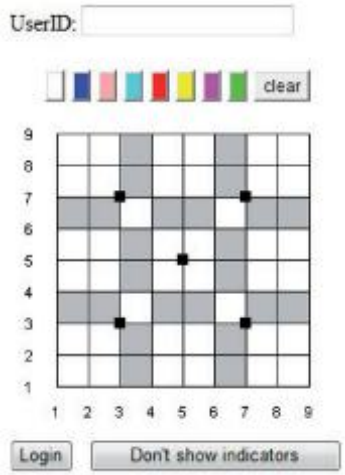
- Recognition Based
- Recall Based
- Cued Recall Based

A part of the research project was initially intended to explore canonical examples of all 3 categories. Although, only the definition of these categories will be discussed as a part of this report.

Recall Based Authentication

Recall-based systems, also known as drawmetric systems, require users to recall and reproduce a secret drawing, typically drawn either on a blank canvas or on a grid. As discussed in the research paper "Graphical Passwords: Learning from the First Twelve Years" by Robert Biddle, Sonia Chiasson, and P.C. Van Oorschot, recall is considered a difficult memory task because retrieval is done without memory prompts or cues. However, users sometimes devise ways of using the interface as a cue, transforming the task into one of cued recall.

Draw a Secret



Pass-Go

Cued-Recall Based Authentication

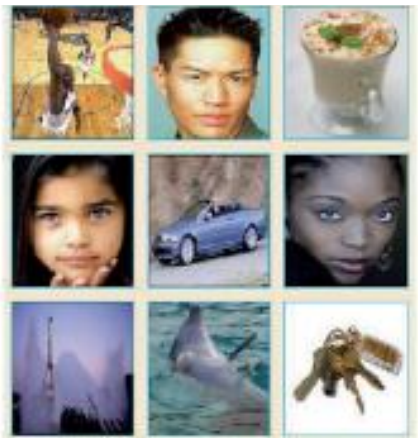
Cued-recall based systems, alternatively referred to as locimetric systems, utilize cues in images to assist users in recalling locations or components within the image for authentication purposes. This approach utilizes cues, striving to strike a balance between security and user-friendliness by ensuring that these cues are beneficial solely to legitimate users. Individuals maintain precise and detailed visual recollections of items they have previously focused on in visual scenes, indicating that users might be capable of recalling specific segments of an image as their password.

Cognitive Authentication Scheme

Recognition Based Authentication

Recognition-based systems, alternatively known as cognometric or searchmetric systems, serve as a mode of authentication where users must identify images they have previously memorized amidst a collection of decoy images to gain access. These systems are especially advantageous in settings where utilizing human cognitive abilities for recognizing images can improve user interaction and possibly augment the security of the login process.

For this study we focus on Passfaces based pin passwords.



Story System

Deja Vu

Vulnerabilities associated with authentication menthods

Weak Passwords:
Using weak passwords that are easily guessable, such as "password," "123456," or common words, poses a significant vulnerability. Attackers can guess these passwords through brute force or dictionary attacks.

Password Reuse:
Reusing the same password across multiple accounts is a common vulnerability. If one account is compromised, attackers can gain access to other accounts where the same password is used.

Lack of Complexity:
Passwords that lack complexity, including a mix of uppercase letters, lowercase letters, numbers, and special characters, are more vulnerable to being cracked.

Short Passwords:
Short passwords are easier to guess or crack through brute force attacks. Longer passwords are generally more secure.

Default Passwords:
Failure to change default passwords on devices or systems can expose them to attacks. Default passwords are often well-known and easily exploited.

Password Recovery Questions:
Weak security questions or publicly available information can be used to reset passwords, making them a vulnerability if attackers can easily guess the answers.

Account Lockout Policies:
Incorrectly configured account lockout policies can either make it too easy for attackers to lock out legitimate users or not provide enough protection against brute force attacks.

No Multi-Factor Authentication (MFA):
The absence of MFA leaves accounts vulnerable to unauthorized access in case passwords are compromised.

Types of Attacks

Brute Force Attack:
In a brute force attack, an attacker tries every possible combination of characters until they find the correct password. This method is time-consuming but can be effective against weak and short passwords.

Dictionary Attack:
In a dictionary attack, attackers use a list of common words or phrases to guess passwords. This approach is more efficient than brute force and can be effective against easily guessable passwords.

Phishing:
Phishing attacks involve tricking users into revealing their passwords by posing as a trustworthy entity, typically through fake websites or deceptive emails.

Keylogging:
Keyloggers are malicious programs or hardware devices that record a user's keystrokes, including passwords, without their knowledge

Rainbow Table Attack:
A rainbow table is a precomputed table of password hashes and their corresponding plaintext passwords. Attackers can use these tables to quickly look up the plaintext password associated with a hash.

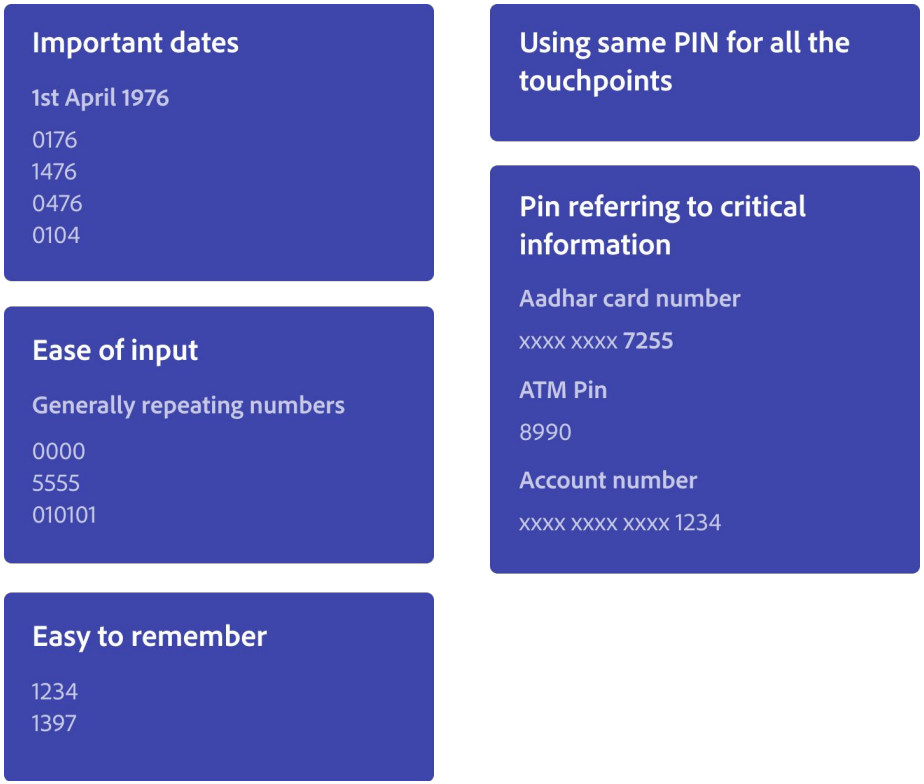
Shoulder Surfing:
Shoulder surfing is a low-tech attack where an attacker physically observes a person entering their password, often in a public place like a café or office.

Shoulder Surfing (Credits: https://octopus-office.co.uk)

Graphical Pin based passwords

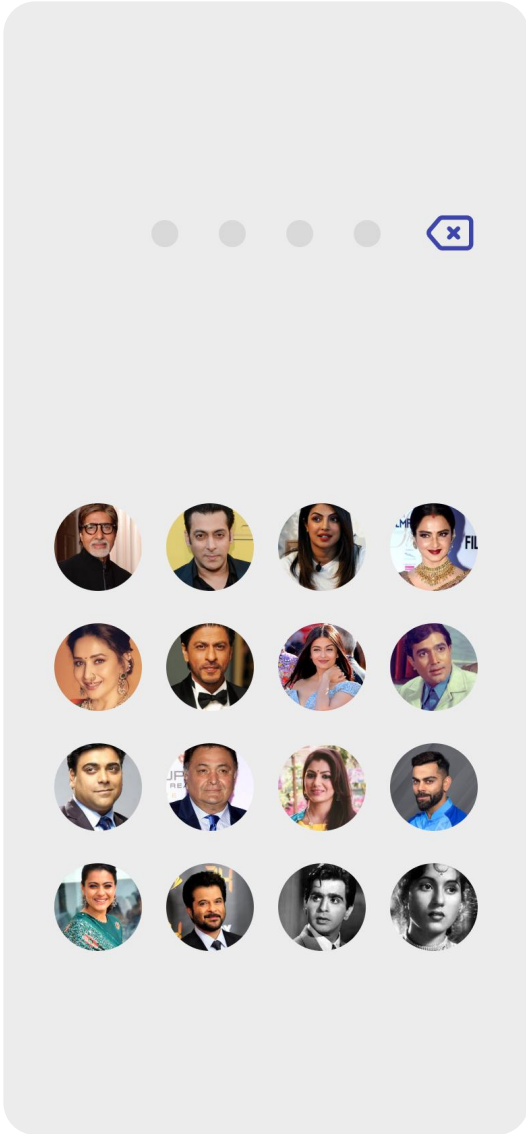
Issues with traditional PIN based passwords

Traditional PINs are a common method of authentication but are often subject to memorability issues. In context of emergent users or those in the older age spectrum, users might find it challenging to remember a string of numbers. The situation is further complicated when users try to mitigate the issue of memorability by opting to use the same PIN across multiple platforms, increasing the risk of security breaches or even using sensitive personal data like adhaar number digits, ATM pins or date of birth.



Why Graphical pin based passwords

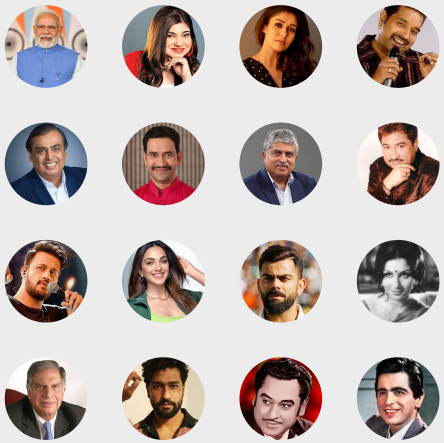
Graphical PIN-based passwords emerge as a promising solution to these challenges. Given the inherent human ability to recall visual elements more effectively than alphanumeric strings, graphical passwords can offer enhanced usability. Users are generally more adept at remembering images, shapes, or patterns, making graphical passwords a more intuitive and user-friendly authentication method. This visual-based approach can potentially reduce the reliance on traditional, more vulnerable PIN systems, thereby enhancing overall security and user experience.



By adopting graphical PIN-based passwords, users can leverage their visual memory to create unique and secure authentication methods, mitigating the risks associated with traditional PINs and contributing to a safer and more user-centric digital environment.

Beyond the security offered by this kind of password, there exists a challenge in systematically developing and assessing a balanced corpus or grid, ensuring each image has an equal likelihood of being selected.

An unbalanced grid



Individuals being significantly more popular/recognizable than others

The probability of them getting chosen in a grid becomes way higher than rest of the elements. This is a potential vulnerability.

Disparity across Regions

Comparatively more recognizable in **Southern India**

Comparatively more recognizable in **North-Central India**

Disparity across Age Groups

Comparatively more recognizable for the **younger generation**

Comparatively more recognizable for the **older generation**

Visual Bias

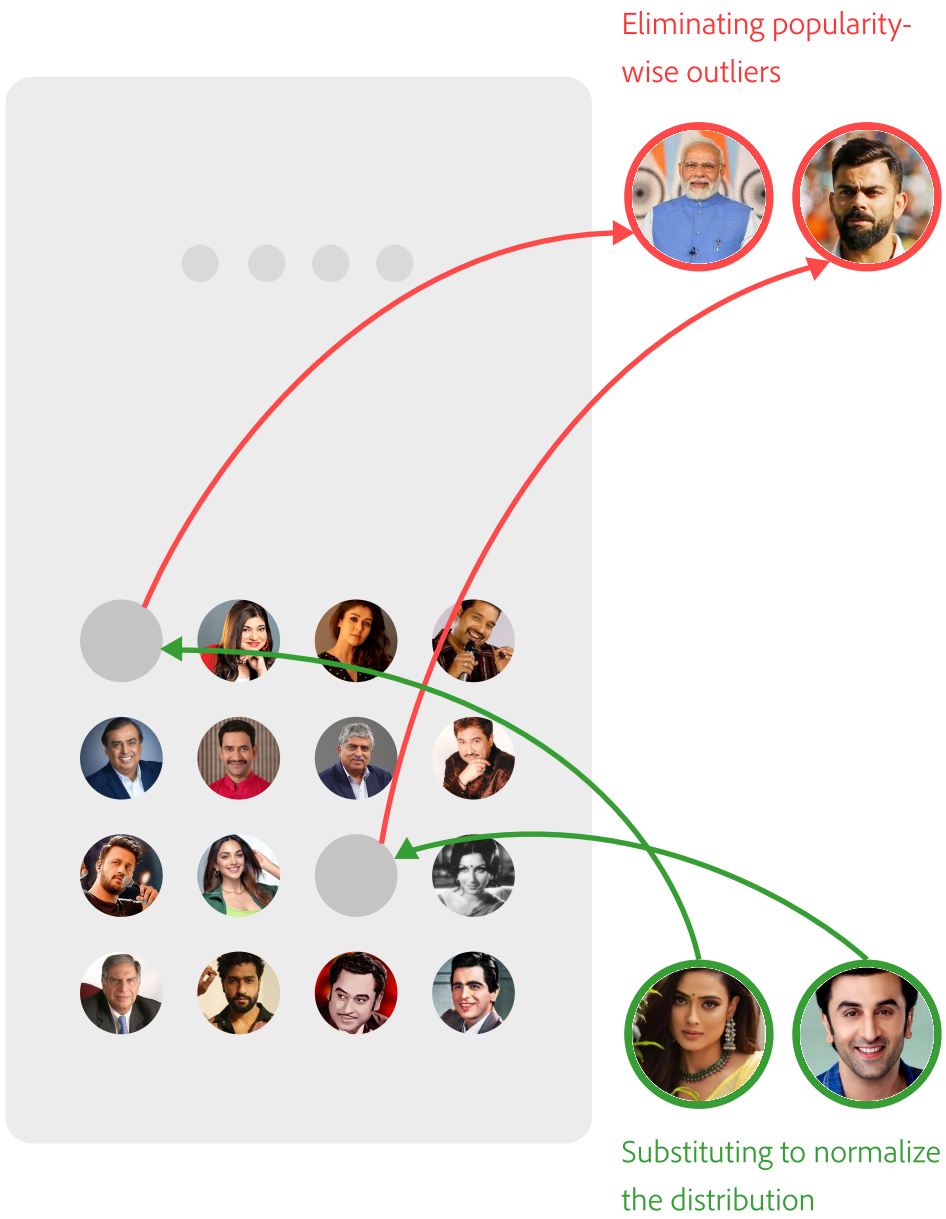
Having unique attires or facial features. Like a turban or a unique haircut

For the purpose of this study we will focus solely on popularity aspect

Why is Grid balancing necessary

Balancing grids in graphical pin-based passwords utilizing celebrity faces is crucial for maintaining the integrity and reliability of the authentication system. An unbalanced grid, characterized by a mix of exceedingly popular and obscure faces, results in predictable password patterns as users tend to choose more recognizable and memorable faces. This predictability increases the vulnerability of the system to unauthorized access and compromises the security of user data.

A balanced grid, where each face has a comparable level of recognition and memorability, mitigates this risk by ensuring that user choices are distributed more uniformly across the available options, thereby reducing the probability of unauthorized individuals successfully guessing the pin. This uniformity is especially vital in regions with diverse celebrity recognition patterns, as it ensures equal representation and avoids biases, making the graphical pins more secure and user-friendly. Balancing also enhances user experience by providing a range of equally recognizable options, facilitating easier recall and faster authentication, crucial for user satisfaction and system efficiency.



Shoulder Surfing in Graphical Passwords

Graphical pin-based passwords offer a unique solution to the pitfalls of traditional alphanumeric passwords by leveraging human memory's affinity for images. However, this innovation is not without its vulnerabilities. One of the significant lingering concerns is the threat of shoulder surfing attacks.

Shoulder surfing occurs when an unauthorized individual intentionally or unintentionally observes or captures a user's authentication process, gaining illicit access. Graphical pin-based passwords, despite their improved memorability and user-friendliness, are particularly susceptible to this threat due to the visually intuitive nature of the authentication elements, enabling easier reproduction of user inputs by onlookers.

Addressing shoulder surfing vulnerability is crucial for maintaining user trust and the overall security integrity of graphical pin-based systems. The visual and intuitive interface, intended to improve user experience and recall, ironically makes observation and replication of the authentication elements more straightforward for malicious entities.

This aspect was initially looked at during exploration of this space. But it wouldn't be investigated more in this study. Although it is a great vertical to dive into for future studies. Nevertheless, some possible design suggestions will be made in this report.

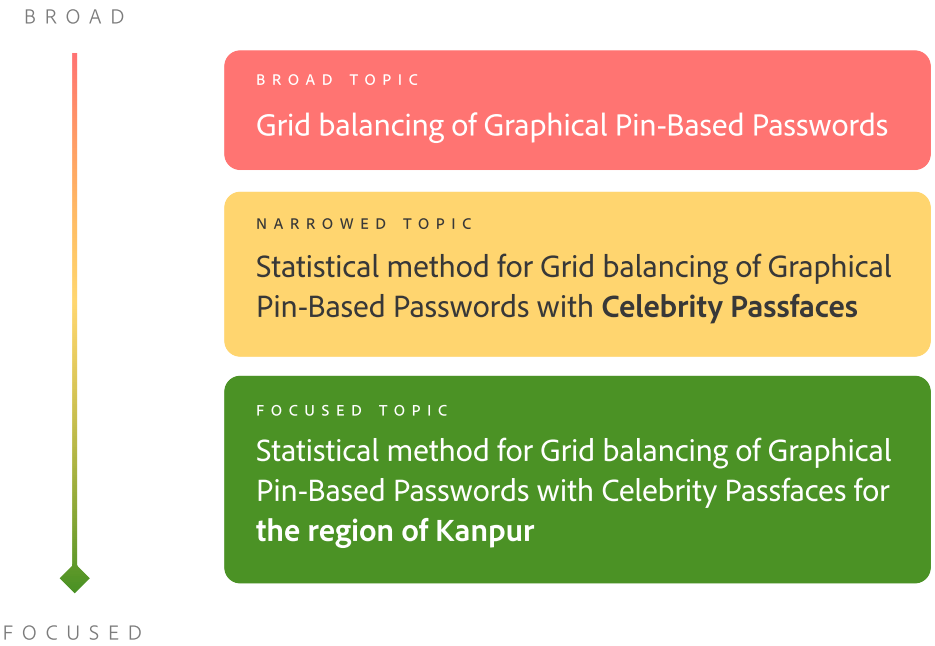
Study details

Research Question

Through this study, 2 questions are intended to be answered.

- How might we evaluate if a grid is balanced or not
- How might we create a balanced grid.

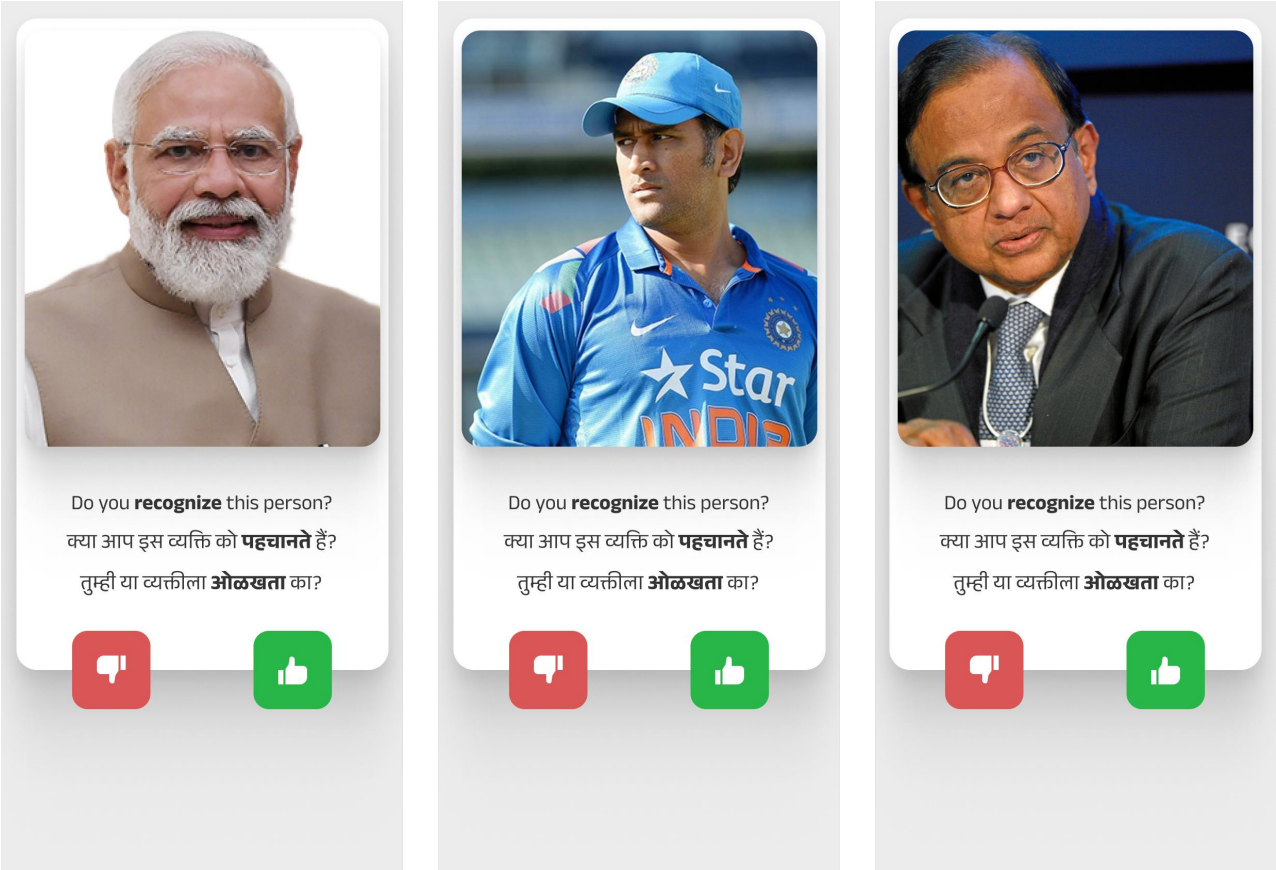
The research question had to be contextualized to make the study feasible.



"Coming up with a **method** to systemically evaluate and create **regionally balanced** grids for **graphical pin-based passwords with celebrity pass faces**."

Ideation for collection of data

An idle yes or no application

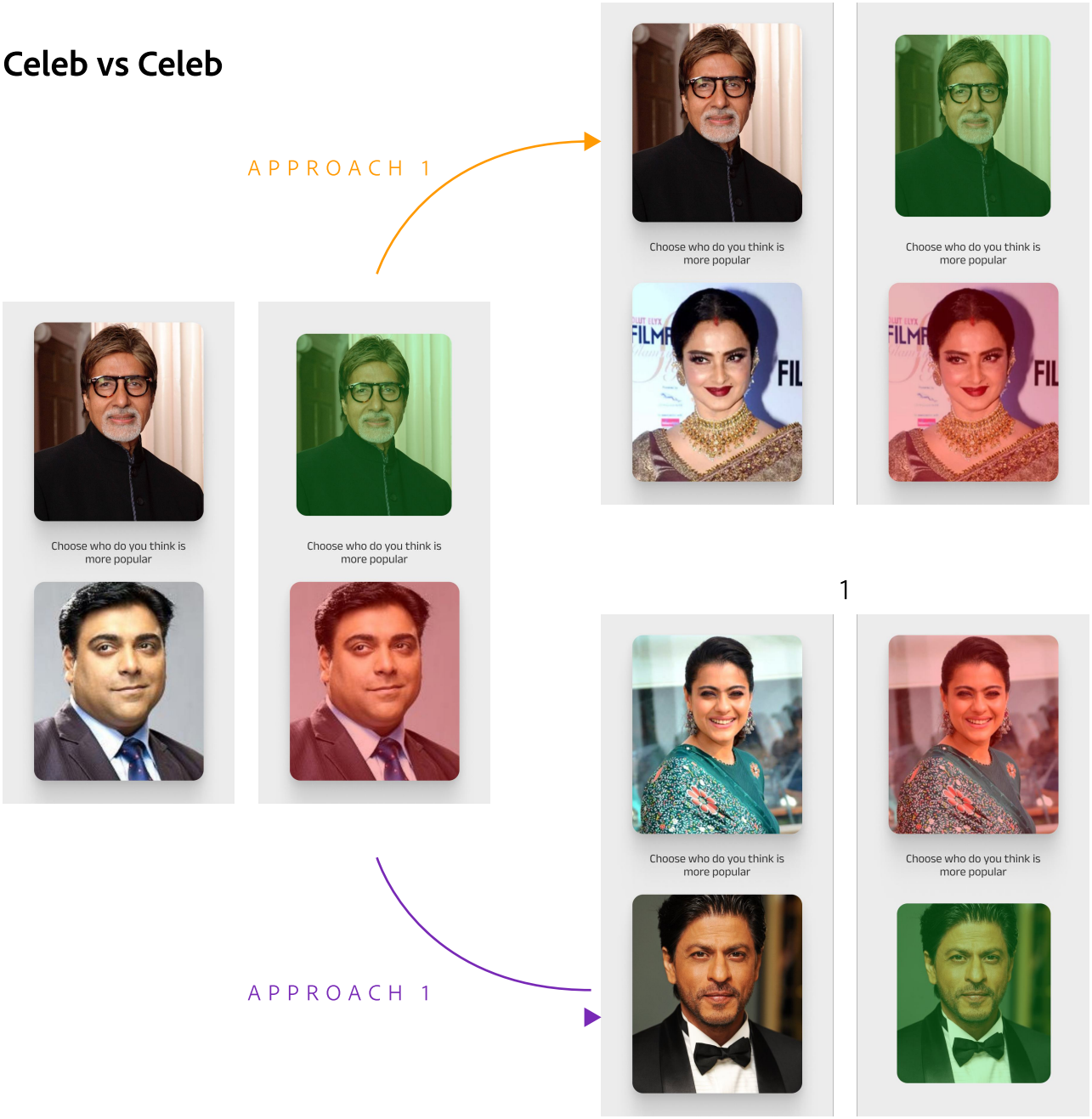


A simple applications where the participants of the study play for specific amount of turns and just answer yes or no to whether they recognize the face or not.

Possible inference

- Recognizability
- Popularity

Celeb vs Celeb



An applications where the celebrities are pitched against each other in a contest of popularity. There are 2 approaches to it.

Approach 1: The celebrity chosen remains on the screen until they are not chosen again.

Possible inference

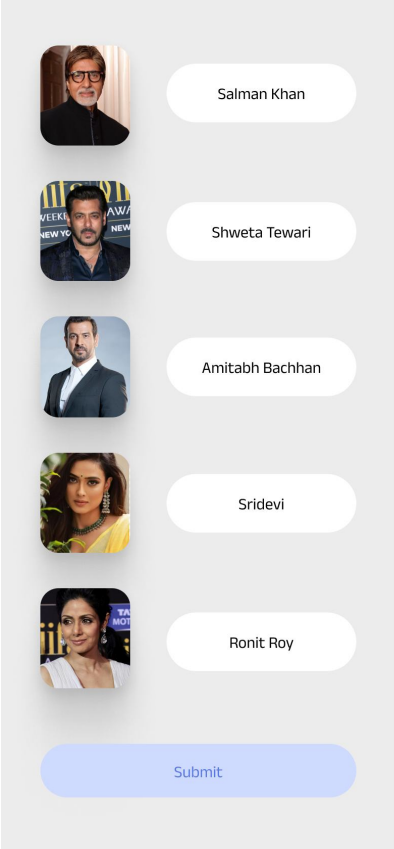
- Comparative popularity

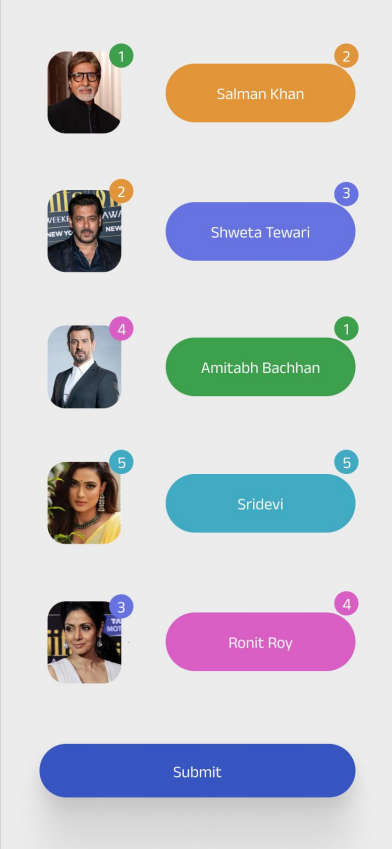
Approach 2: The pair changes every round

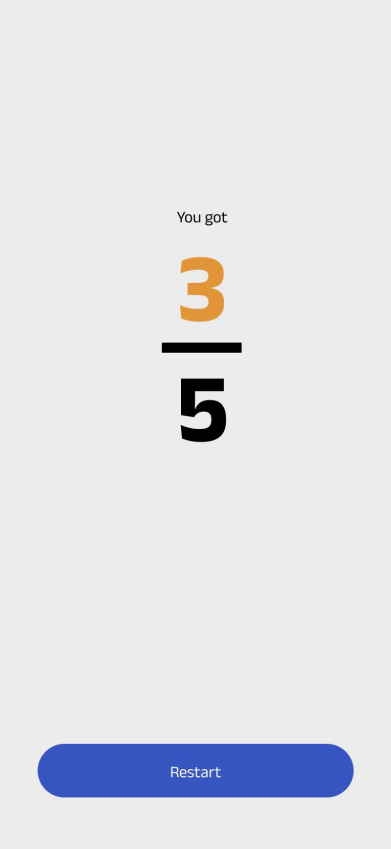
Possible inference

- Popularity

Match the following





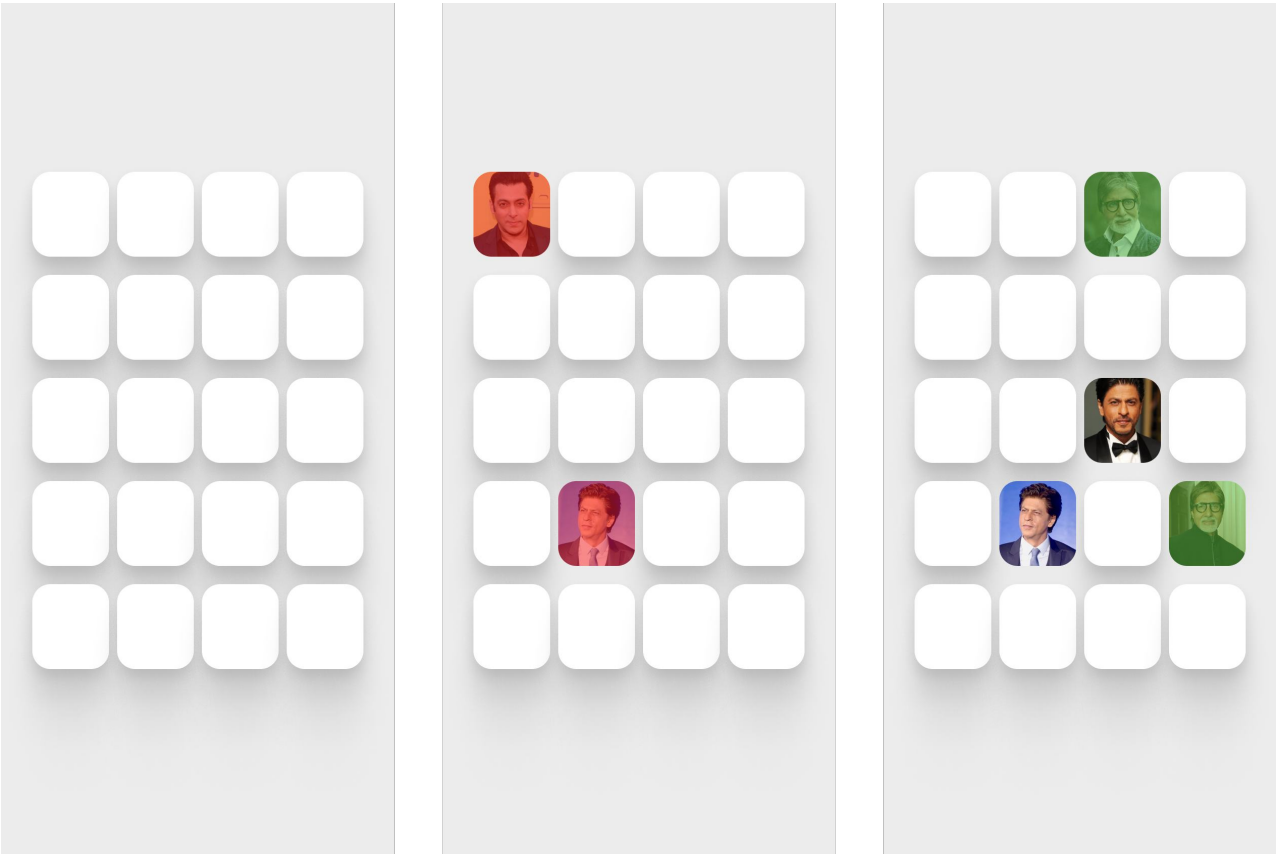


A simple applications where the participants of the study play for specific amount of turns and just answer yes or no to whether they recognize the face or not.

Possible inference

- Recognizability
- Popularity

Flip-Card Game

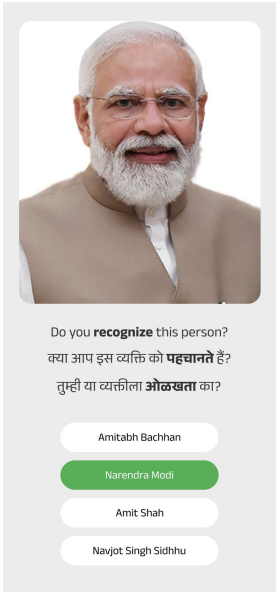


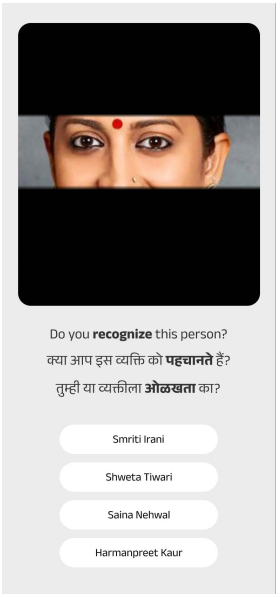
A game where the user has to find and pair 2 images of celebrity consecutively. The goal is to open up all the cards.

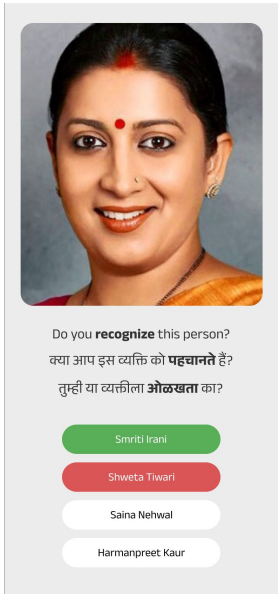
Possible inference

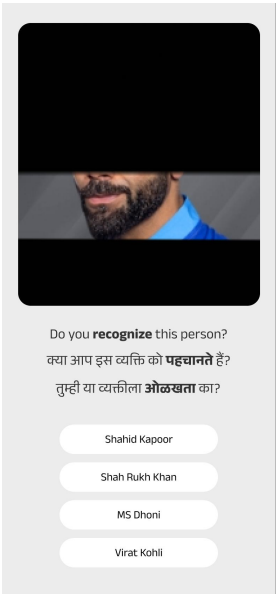
- Recognizability
- Can also take in account the feature based or visual recognizability

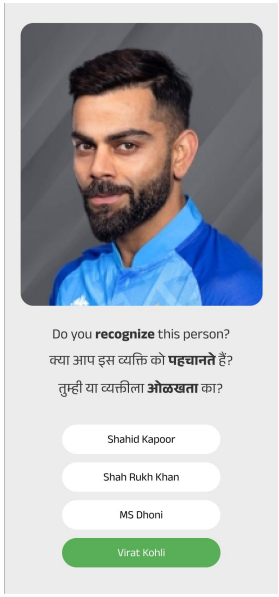
Using partially visible images for identification











A questionaire where users have to answer with the name of the celebrity they recognize by partially obsured image of them, such that a feature is visible.

Possible inference

- Recognizability through features

Issue with the ideation

On assessment of these ideas, a problem which surfaced was how are we going to fixate on a corpus for a region. All the ideas required a fixed corpus to function on, which initially was thought to be kept arbitrary or as per the taste of the researcher. But this neglects regional biases and adds our own to it.

Hence we had to come up with a systematic way where we can concretize a corpus or avoiding a corpus altogether for gauging popularity and recognizability of the individual.

Initial data collection approach

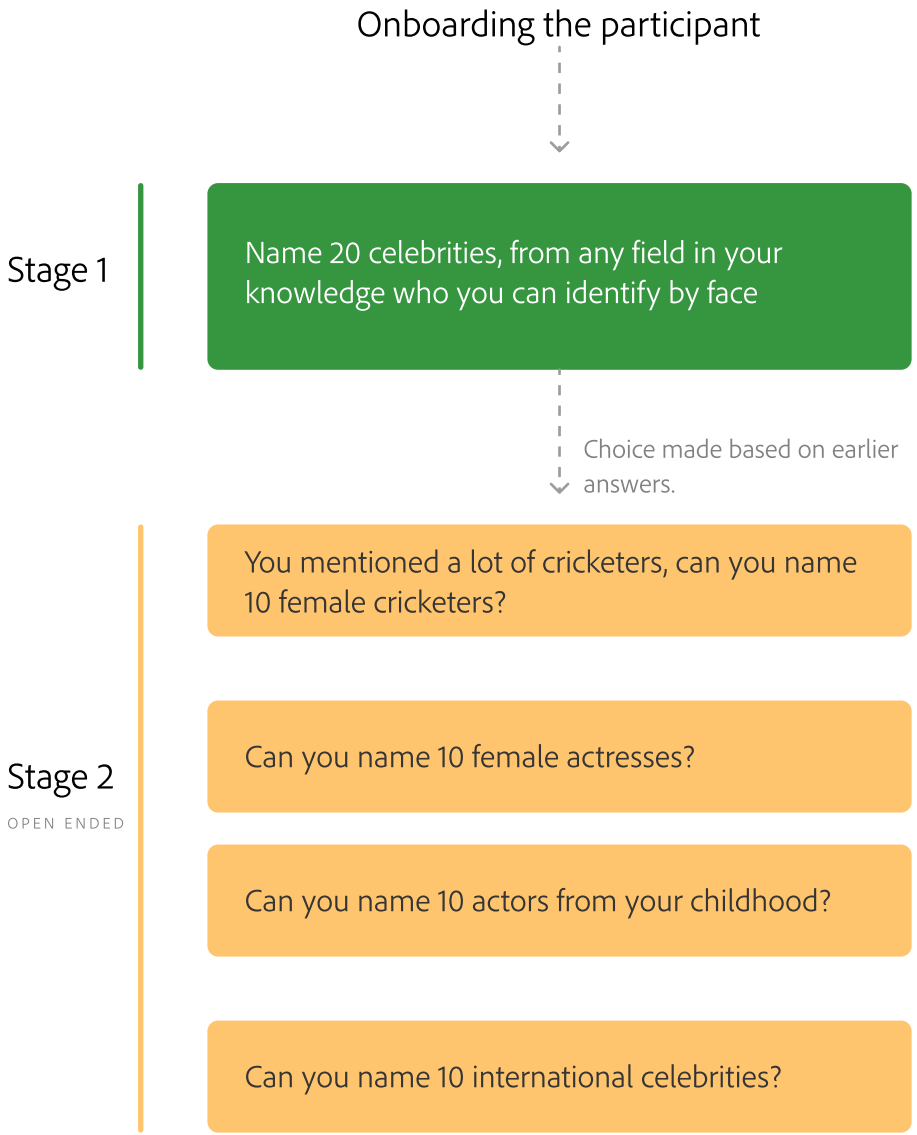
Protocol

An essential consideration to be made was how we are probing users. The plan for pilot was to get **50** names from each participant through questioning. A choice had to be made whether to keep the engagement open ended or controlled, where we tried to align to an open ended approach.

The questionnaire was basically classified in two stages. The goal was to ask 1 question in stage 1 and 3-4 questions in stage 2. The target for number of names we get from each questions were:

Question 1: 20
Question 2 and after: 10-5

The sequence is described in the diagram



Plan to come up with a evaluation method

Envision

The vision is to calculate a **Popularity index (P_i)** for celebrity X, on the basis of 2 factors:

- **M_T:** How many times was the name mentioned at every stage of task.

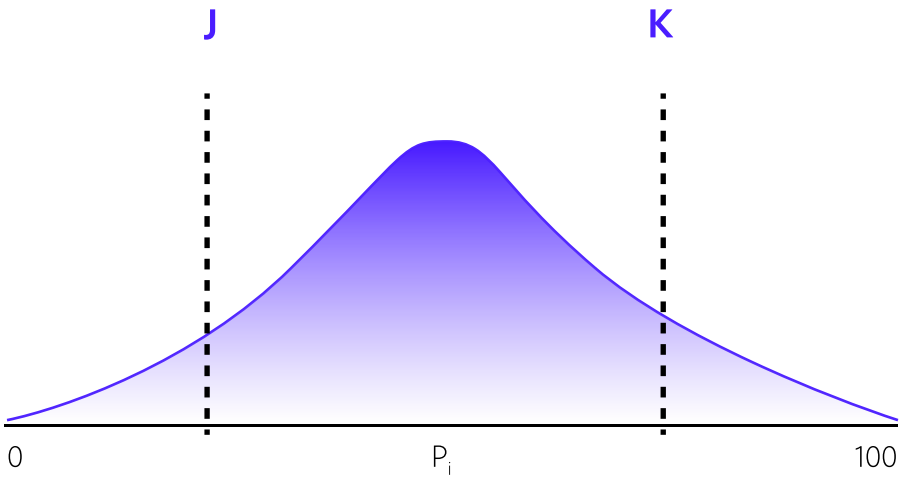
  M = M_1= Frequency mentioned in first round
  M = M_2= Frequency mentioned in second round

- **n_T:** Was the name mentioned after the first question or on further probing. It will be a coefficient.

  N = N_1, if mentioned in first round,
  N = N_2, if mentioned in second round

$$P_i = M_1 * N_1 + M_2 * N_2$$

0<Pi<100



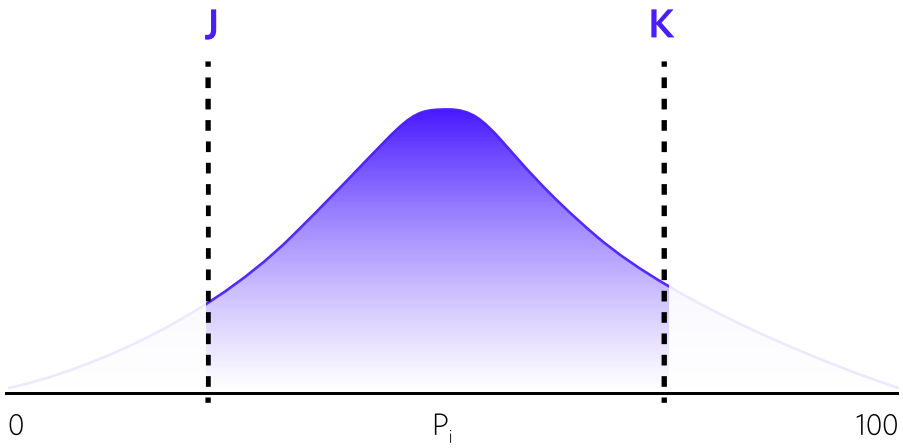
We might be able to get a normal distribution of P_i . Since our goal is to identify outliers and remove it. We come up with 2 bounding variables.

K: P_i at x %ile of the sample
(x > 50%ile, will decide on a value and test)

J: P_i at y %ile of the sample
(y < 50%ile, will decide on a value and test)

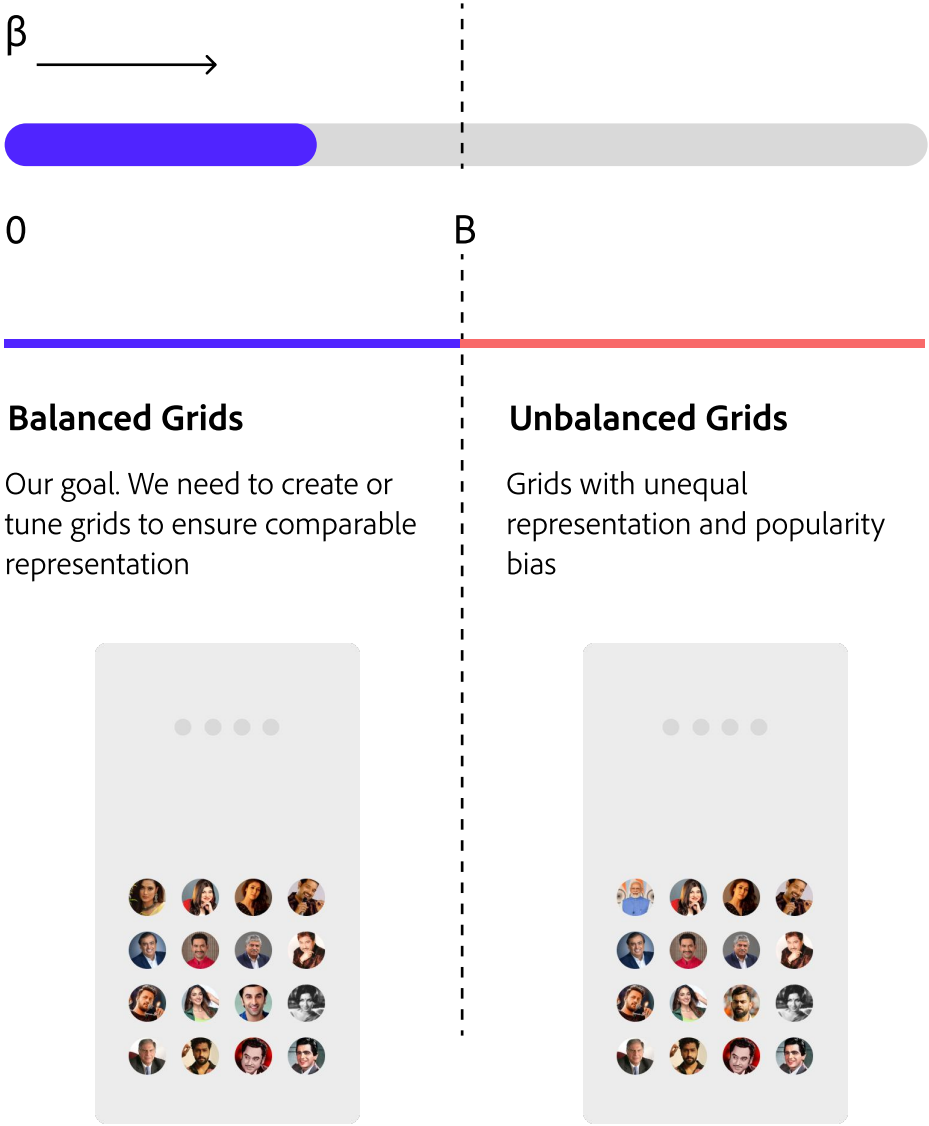
We will arrive at **Balance index (β)** of the grid.

$$\beta = | K - J |$$



Finally we declare the **Balance constant B.** B would be decided and tuned after feedback till it gets stagnant.

B if β > B, Grid is **unbalanced**.
  if β < B, Grid is **balanced**.



1st Pilot Deployment

The implementation was carried out for four users, involving real-time data input and phone-based activity recording. Generally, the protocol was adhered to, with minor modifications in each iteration.
Prominent figures such as Narendra Modi and Virat Kohli frequently appeared in the initial stage for all cases.
However, throughout the exercise, a common phrase cited was, "I recall the face but can't recall the name."
Achieving the target of 50 names in the activity proved challenging, as the average number of names participants could remember was 32 before concluding the exercise.

Learnings

- Recognizing a name doesn't necessarily imply the ability to recognize a face, and vice versa. Therefore, it was critical to include both recognition and recall processes in this research. The current methodology primarily assesses the recall capabilities of the participants.
- To guarantee a diverse range of data, the study should be conducted with a larger sample size. For establishing the constants and limits outlined in the evaluation section, ongoing experiments with varying parameters are required to identify a stable value that could serve as a standard for subsequent research.
- Furthermore, to gather the requisite amount of data, participants need to be involved in the study for a longer time period per session.

Recognition vs Recall

Recognition entails identifying previously learned information when it is encountered again, whereas recall involves retrieving information from memory without any explicit prompts.

Regarding graphical passwords, users are typically offered a selection of images (in this case, celebrities) from a larger set during the setup phase, where they **recognize** and choose their preferred celebrities. During login, users are shown a grid that includes both their selected images and additional decoy images. Here, they must recognize and select the images they initially chose.

However, **recall** plays a role during the login process as well. Users must not only remember the images they chose during the setup phase but also the order in which they selected them.

In reference to 1st Pilot

We requested users to name celebrities whom they could identify by face, an activity that primarily tests the ability to recall names. However, this approach overlooked the aspect of recognition. As previously mentioned, recognition and recall are both integral components in the context of graphical passwords.

Consequently, it became necessary to develop a protocol that takes into account both the recallability and recognizability of the celebrities. This dual-focus approach ensures a more comprehensive assessment of user interaction with graphical password systems.

Curating the protocol

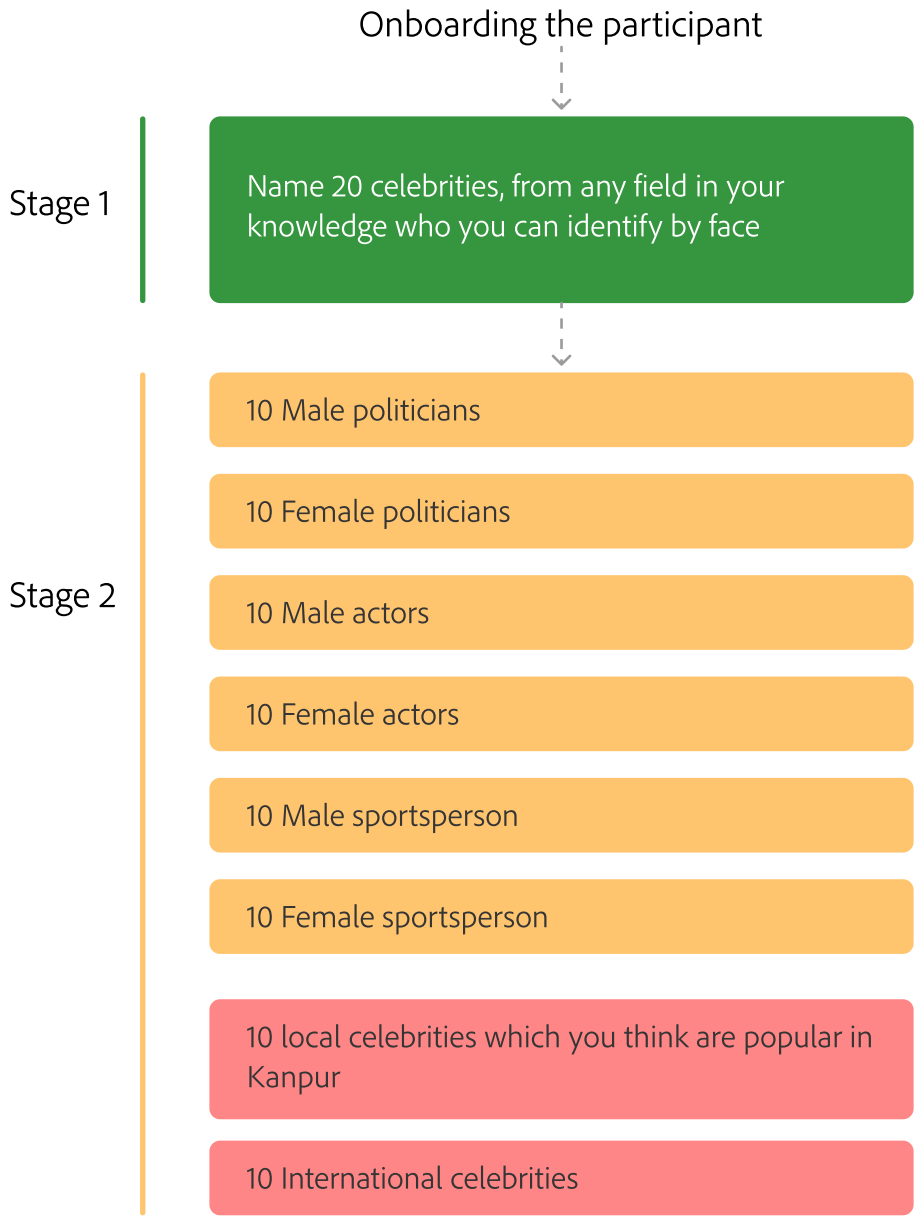
Step 1

Largely remains unchanged where we ask the user to give 20 names of celebrities from any field they can recognise by face.

Step 2

Opting for a more structured methodology, I chose to employ fixed questions to streamline the data collection process. Certain categories seemed disproportionately represented in the pilot. While this approach has its constraints, I addressed this in the protocol by including questions that were not confined to any particular field of work, focusing instead on pure recall. A total of eight questions were posed, details of which can be found in the accompanying diagram.

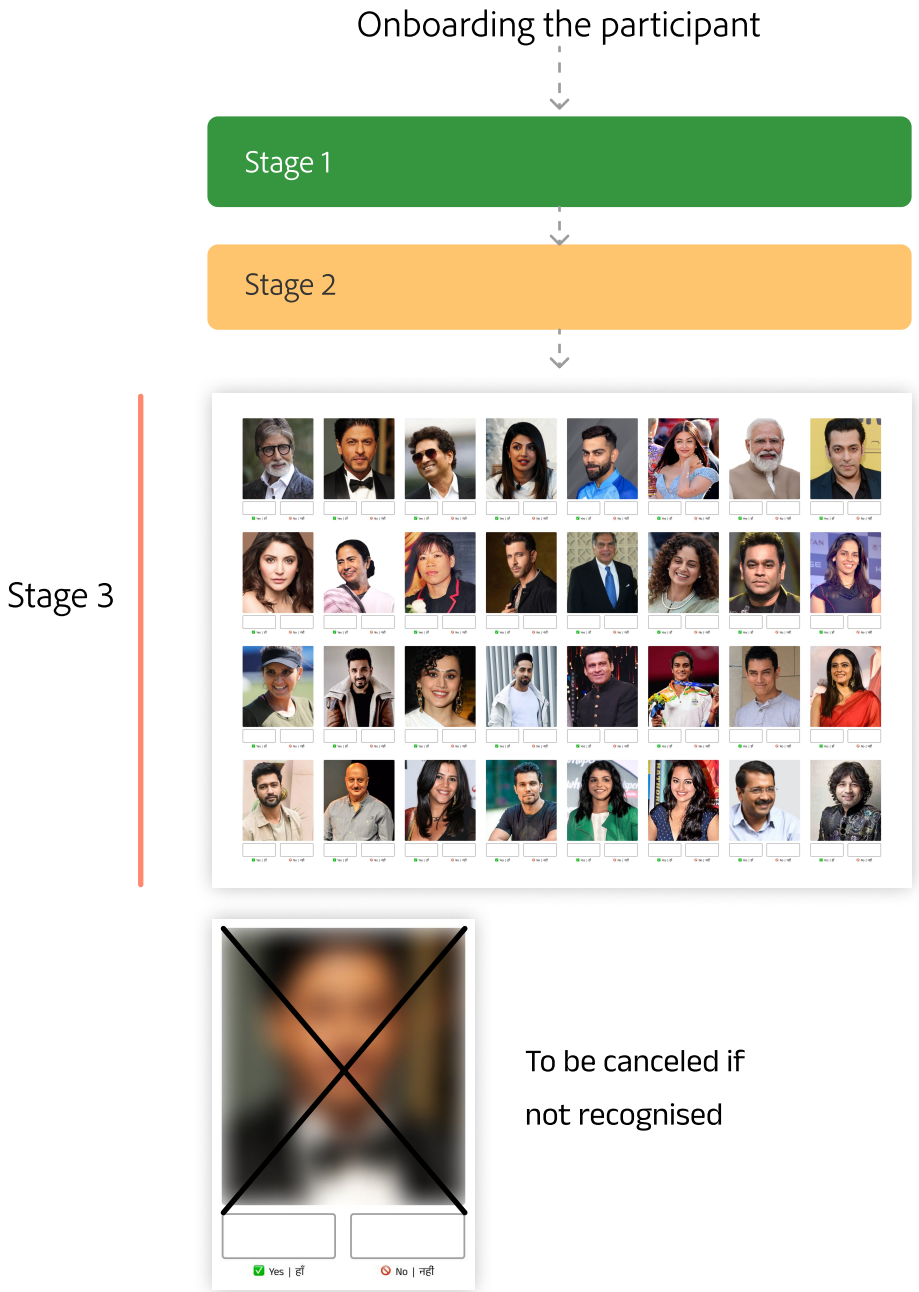
Question 1: 20
Question 2-8: 10 (atmost)
Total: 100



Step 3

This phase concentrates on the recognizability of celebrity faces. Initially, I compiled a list of 200 Indian celebrities, creating a diverse corpus by researching the most popular figures across various fields.

Following this, participants were provided with a sheet displaying the faces of these celebrities. The instruction given was to mark off any celebrity face that was not recognizable to the participant. It was emphasized that not knowing the name was acceptable, as long as the participant could recognize the celebrity's face. This task was designed to assess the ability to recognize faces irrespective of the ability to recall names.



2nd Pilot Deployement

The implementation was carried out with 10 users in Kanpur, involving real-time data entry and recording of the activity over the phone. Each user took approximately 45 minutes to complete the activity. Due to the extended duration of the study, participants were able to identify an average of 80 names, with the option to skip questions if they couldn't recall any further names. Consequently, the outcomes for stages 1 and 2 mirrored those of the first pilot.

Stage 3 yielded intriguing results. Notably, several celebrities mentioned in the earlier stages were not part of the pre-arranged corpus. Additionally, on average, participants were unable to recognize about 10 faces from the corpus, highlighting interesting insights into celebrity recognizability among the participants.

Learnings

In addition to the insights gained from the first pilot, the second pilot revealed several key findings:\

- Participants tended to name celebrities in a particular sequence, largely influenced by their affiliations and fields. This trend was more pronounced in the open-ended questions of stages 1 and 2. For instance, upon mentioning the well-known Indian cricketer Virat Kohli, the subsequent names were often those of other cricketers.
- It became evident that the Stage 3 corpus needs to be dynamically adjusted. This involves removing celebrities who are frequently unrecognized and replacing them with those most often recognized in stages 1 and 2. This adjustment should be based on a "Recognition Index (R_i)," which will be elaborated on later, to ensure the corpus remains relevant and reflective of the participants' recognition abilities.

Finalized Method

The vision is to calculate a **Recognizability index (R_i)** for celebrity X, on the basis of 2 factors:

- **M_T:** %age of users mentioning the name in any stage

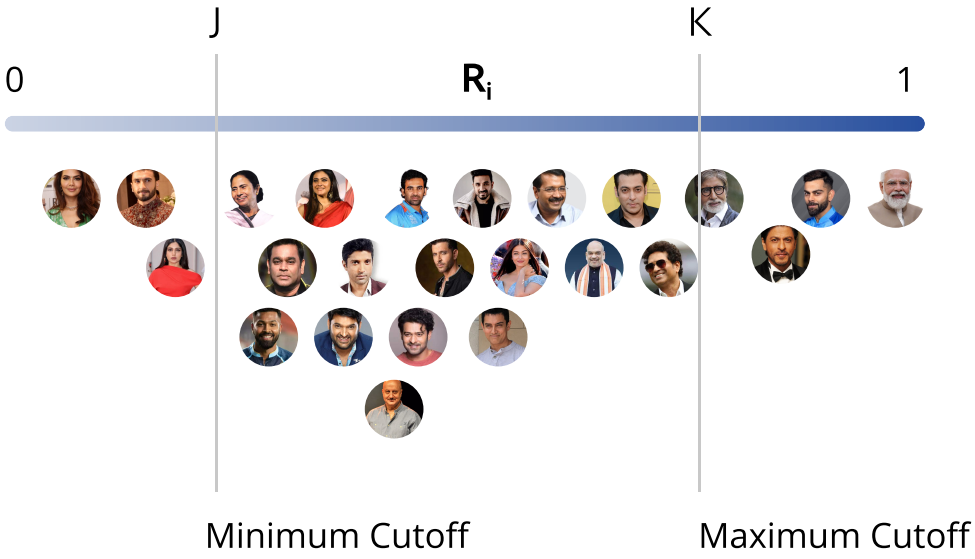
  M = M_1 = Proportions of users in 1st Stage
  M = M_2 = Proportions of users in 2nd Stage
  M = M_3 = Proportions of users in 3rd Stage

- **N_T:** Was the name mentioned after the first question or on further probing. It will be a coefficient.

  N = N_1 , if mentioned in first round, = **0.5**
  N = N_2 , if mentioned in second round = **0.3**
  N = N_3 , if mentioned in third round = **0.3**

$$R_i = M_1*0.5 + M_2*0.3 + M_3*0.3$$

Where 0<Ri<1

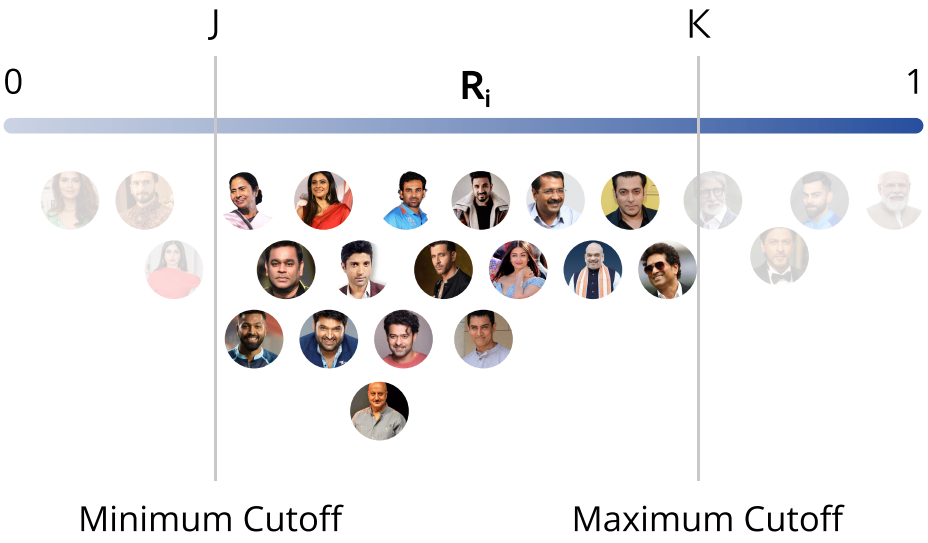


We might be able to get a distribution of R_i . Since our goal is to identify outliers and remove it. We come up with 2 bounding variables.

K: Top x % of distribution = 5%

J: Bottom x% of distribution = 50%

Note: Reason to keep the bottom trim high is that I hypothesise that the distribution will be left-skewed due to diversity, ie: a lot of names with low R_i.

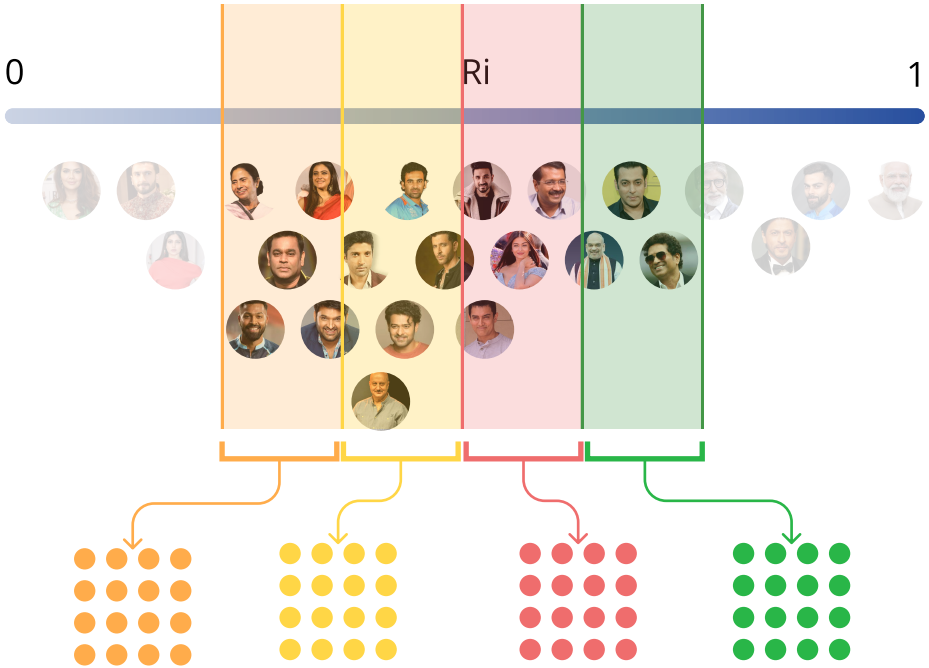


How to balance a corpus

We can eliminate the celebrities with R_i above maximum and below minimum cutoff, to eliminate the outliers in popularity. Hypothetically we will be left with a set of corpus which has a relatively low disparity in popularity and henceforth a similar probability of getting selected.

Note: The grid can never be entirely balanced in real world scenarios, ie likelihood of every photo of being selected can never be equal.

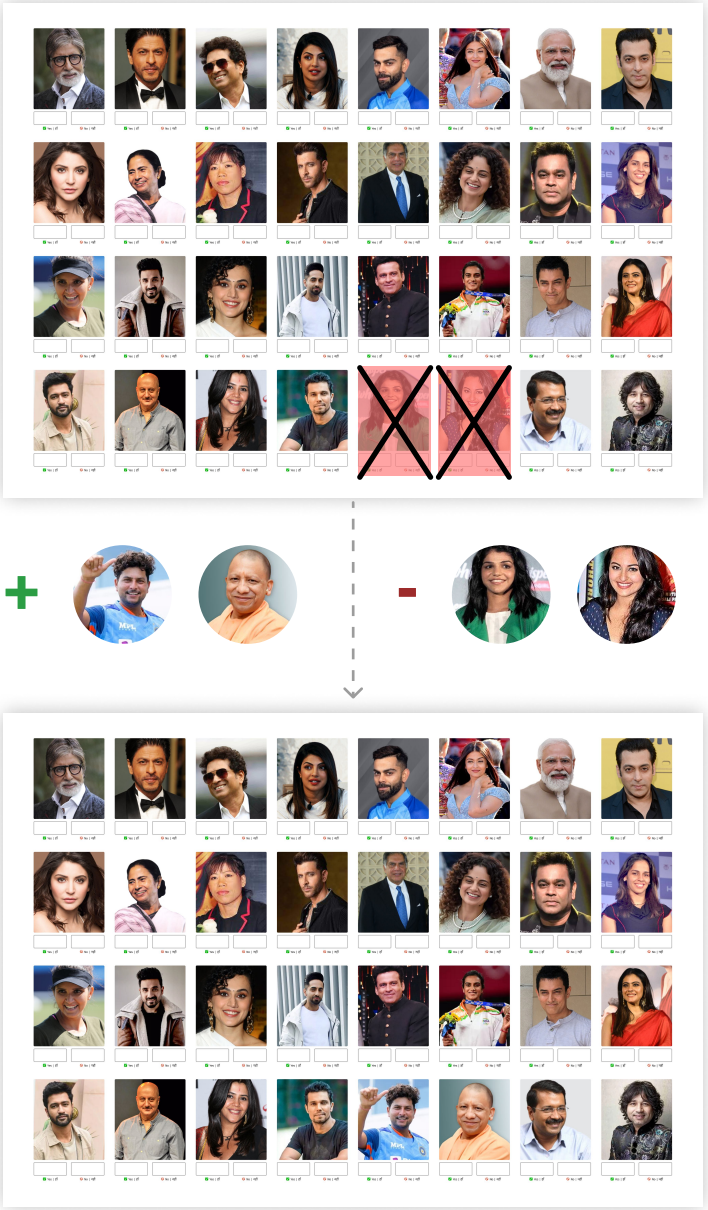
Constructing balanced grids



The plan is to divide the left over distribution into 4 partitions, such as to make one grid out of each partition. This will make sure that there is a lower disparity in popularity across the grid, as the difference in R_i across a partition would be lesser compared to the whole set.

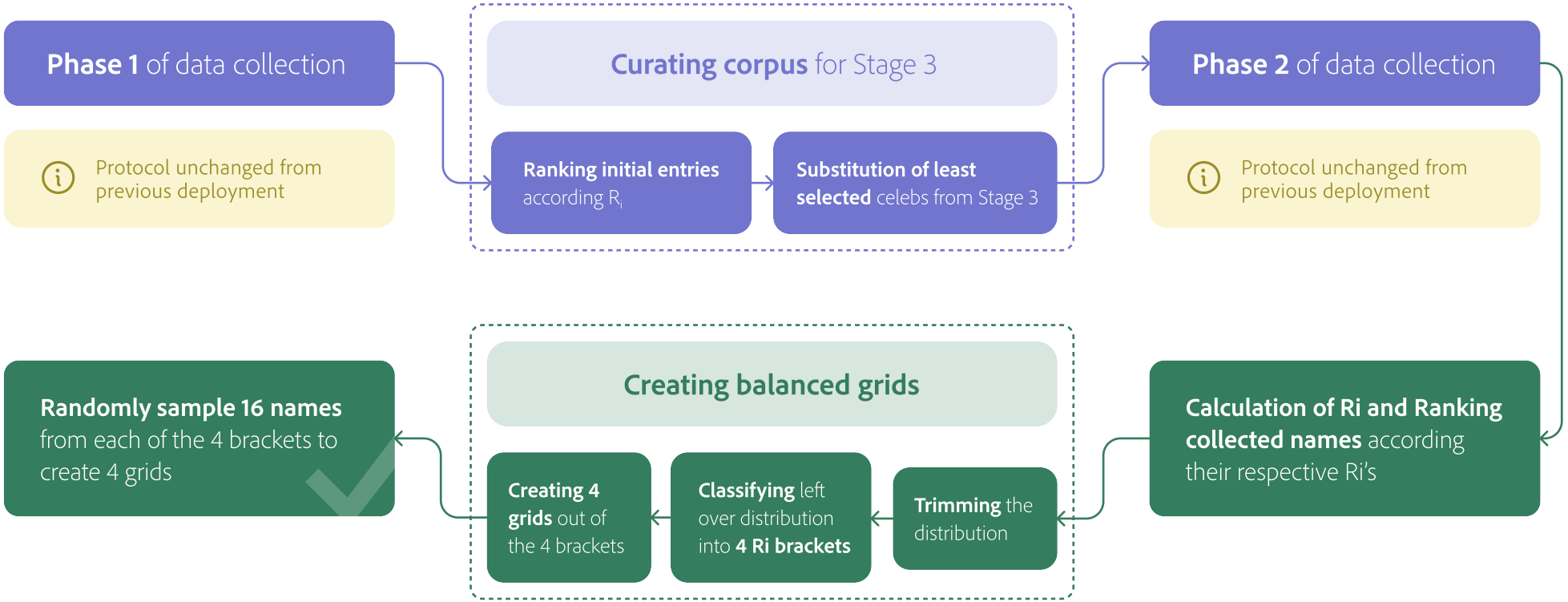
Continuous Adaptability

As pointed in the 2nd pilot, another addition to the study can include use updating the corpus of stage 3 after every participant or set of participants. This ensures updation of corpus to be more locally relevant and the evaluation of R_i more balanced.



Deployment Plan

For the final deployment, I decided to follow the workflow given below. Keeping the protocol majorly the same and integrating a method for continuous adaptability of grids. This was followed by calculation Ri for collected names, subsequent trimming of distribution, and implementing a systematic logic for creation of balanced grids. Finally these grids were evaluated against the grids used in Zuha Asif's study on picture pass faces.

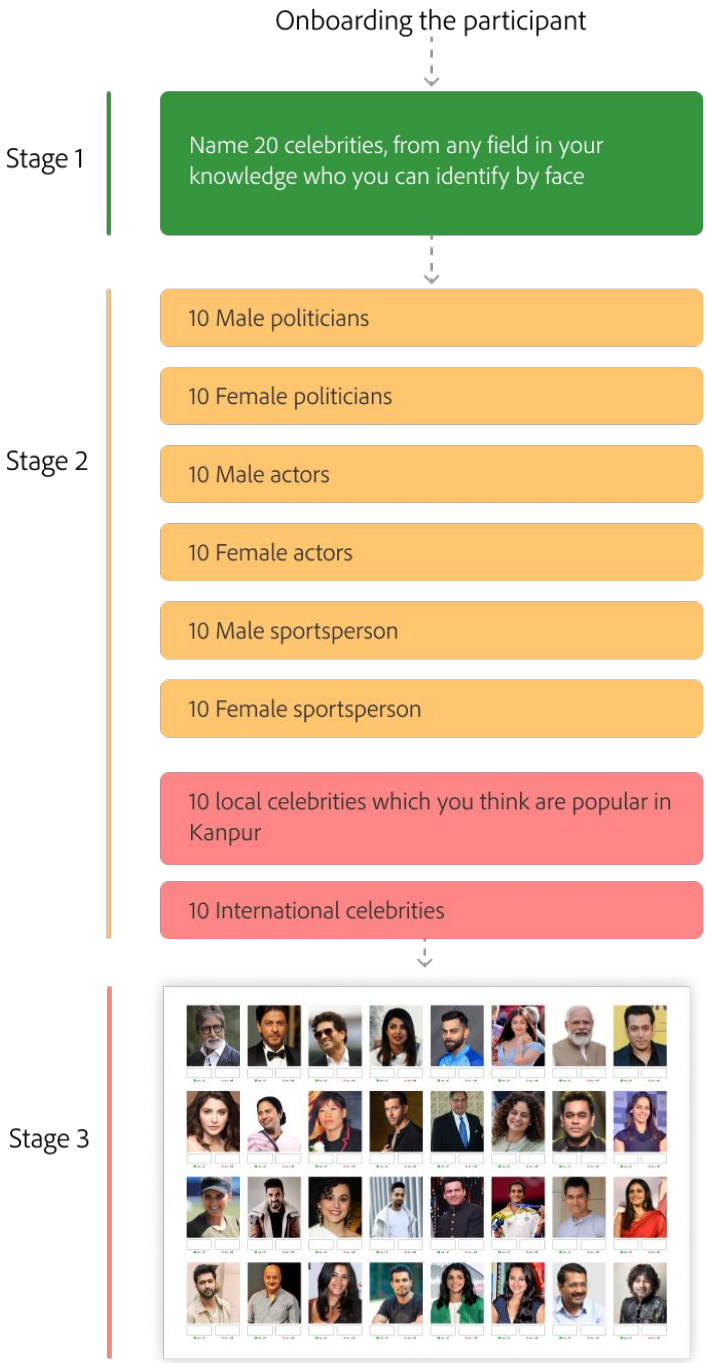


Final Deployment

Phase 1

For the data collection of phase 1, 3 users were onboarded in initially. The protocol was kept the same as the 2nd Pilot. The intention of this phase was to employ the previously described method for "continuous adaptability of the corpus". By the end of this phase, we had close to 330 unique names of celebrities with their calculated Ri. Below is a glance at the data at the end of this phase.

Name	% Stage 1	% Stage 2	% Stage 3	Ri
A.B Diviliers	0	0.3	0	0.3
Aamir Khan	0.5	0.3	0.9	1.7
Aishwarya Rai	0.5	0	0.9	1.4
Ajay Devgn	0.5	0.3	0	0.8
Ajay Jadeja	0	0.3	0	0.3
Akhilesh Yadav	0.5	0	0	0.5
Akshay Khanna	0	0	0.9	0.9
Akshay Kumar	1	0	0.9	1.9
Alia Bhatt	0	0	0.9	0.9
Alok Mishra	0	0.6	0	0.6
Amir Khan	0.5	0	0	0.5
Amit Mishra	0	0.3	0	0.3
Amit Shah	1.5	0	0.9	2.4
Amitabh Bachch	0.5	0.3	0.9	1.7



Continuous Adaptability of the corpus.

Now we had to identify the faces which have been consistently not recognised in Stage 3. For this we substitute them with names with highest Ri's.

Identify ***n*** names with Stage 3 recognition %age < 30%

Substitute with ***n*** names, with highest Ri which were not part of the corpus

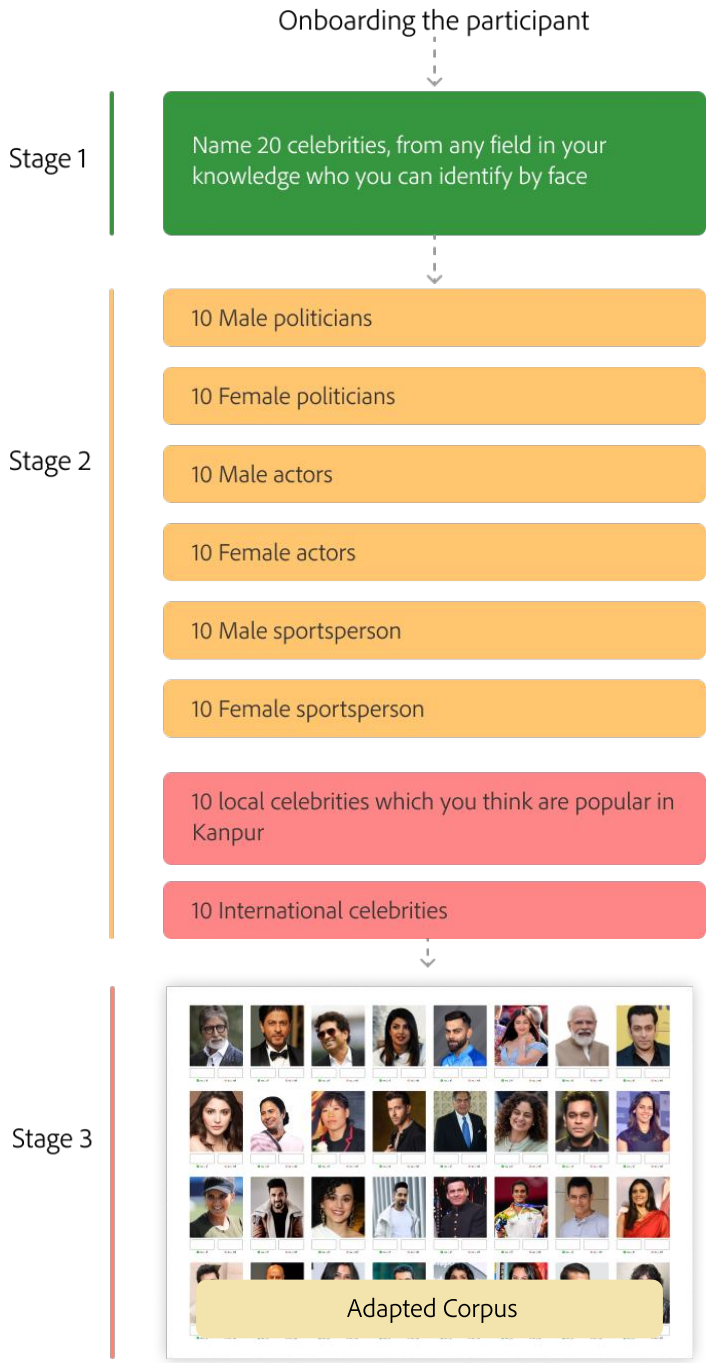
To Eliminate			To Add	
Kiran Rao	0		Hema Malini	0.9
Rajeev Chandra:	0		Jitendra	0.9
Gagan Narang	0		Mulayam Singh Yadav	0.9
Geeta Phogat	0		Ajay Devgn	0.8
Anand Mahindra	0.3		Yogi Adityanath	0.8
Rajyavardhan Si	0.3		Alok Mishra	0.6
Piyush Goyal	0.3		Dharmendra	0.6
Kajal Aggarwal	0.3		Dimple Kapadia	0.6
Atul Gawande	0.3	→	Kapil Dev	0.6
Randeep Surjew	0.3		Kuldeep Yadav	0.6
Shriya Saran	0.3		Mayawati	0.6
Kailash Satyarth	0.3		Nargis	0.6
Nandita Das	0.3		Narsimha Rao	0.6
Radhika Madan	0.3		Neelima Katiyar	0.6
Swara Bhasker	0.3		PT Usha	0.6
Vijay Deverakon	0.3		Akhilesh Yadav	0.5
Geet Sethi	0.3		Raju Srivastava	0.6
Vani Kapoor	0.3		Satish Mahana	0.6
Gurpreet Singh S	0.3		Satyadev Pachori	0.6
Gauri Shinde	0.3		Sunil Gavaskar	0.6

P1 Corpus Score		
Name	Stage 3%	Ri
Kiran Rao	0	0
Rajeev Chandrasekhar	0	0
Gagan Narang	0	0
Geeta Phogat	0	0
Anand Mahindra	0.3	0.3
Rajyavardhan Singh Rath	0.3	0.3
Piyush Goyal	0.3	0.3
Kajal Aggarwal	0.3	0.3
Atul Gawande	0.3	0.3
Randeep Surjewala	0.3	0.3
Shriya Saran	0.3	0.3
Kailash Satyarthi	0.3	0.3
Nandita Das	0.3	0.3
Radhika Madan	0.3	0.3
Swara Bhasker	0.3	0.3
Vijay Deverakonda	0.3	0.3
Geet Sethi	0.3	0.3
Vani Kapoor	0.3	0.3
Gurpreet Singh Sandhu	0.3	0.3
Gauri Shinde	0.3	0.3
Vir Das	0.6	0.6
Manoj Bajpayee	0.6	0.6

Phase 2

For the data collection of phase 2, 9 more users were onboarded. The protocol was kept the same as the 2nd Pilot. The intention of this phase was to get a diverse set of names with the adapted corpus in play.

By the end of this phase we had 524 unique celebrity names.



Eliminating duplicates

Due to the nature of data collection, names occurred across all the stages multiple times. Hence first we had to create a master set of names, which was the union of all unique names mentioned across all the stages.

Stage 1 Unique Entries	Stage 2 Unique Entries	Stage 3 Unique Entries	Master Set
Rohit Sharma	Imran Khan	Amitabh Bachchan	Rohit Sharma
Ms Dhoni	Ranbir Singh	Shah Rukh Khan	Ms Dhoni
Virat Kohli	Sushant Singh Rajput	Sachin Tendulkar	Virat Kohli
Narendra Modi	Sunny Deol	Priyanka Chopra	Narendra Modi
Amit Shah	Bobby Deol	Virat Kohli	Amit Shah
Arvind Kejriwal	Dharmendra	Aishwarya Rai	Arvind Kejriwal
Anushka Sharma	Jitendra	Narendra Modi	Anushka Sharma
Kareena Kapoor	Tushar Kapoor	Salman Khan	Kareena Kapoor
Shah Rukh Khan	Ritesh Deshmukh	Deepika Padukone	Shah Rukh Khan
Salman Khan	Arjun Kapoor	MS Dhoni	Salman Khan
Amir Khan	Shradhha Kapoor	Rajinikanth	Amir Khan
Amitabh Bachan	Illeana De Cruz	Hrithik Roshan	Amitabh Bachan
Rishi Kapoor	Isha Gupta	Akshay Kumar	Rishi Kapoor
Ranbir Kapoor	Kirti Sanon	Shahid Kapoor	Ranbir Kapoor
Saif Ali Khan	Bhumi Pednekar	Kareena Kapoor	Saif Ali Khan
Akshay Kumar	Rashmika MAndhana	Alia Bhatt	Akshay Kumar
Ajay Devgn	Rakul Preet Singh	Anushka Sharma	Ajay Devgn
Priti Zinta	Tabu	Ranveer Singh	Priti Zinta
Rani Mukherji	Rekha	Mamata Banerjee	Rani Mukherji
PV Sindhu	Hema Malini	Mary Kom	PV Sindhu
Madhuri Dixit	Jasprit Bumrah	Ratan Tata	Madhuri Dixit
Sridevi	Hardik Pandya	Kangana Ranaut	Sridevi
Amitabh Bachchan	Kurnal Pandya	AR Rahman	Amitabh Bachchan
Aishwarya Rai	Bhuvneshwar Kumar	Saina Nehwal	Aishwarya Rai
Aamir Khan	Mohit Sharma	Kapil Sharma	Aamir Khan
Nana Patekar	Arshdeep Singh	Rohit Sharma	Nana Patekar
Emraan Hashmi	Ashok Dinda	Priyanka Gandhi Vadra	Emraan Hashmi
Deepika Padukone	Amit Mishra	Karan Johar	Deepika Padukone
Anupam Kher	Vinod Kambli	Yuvraj Singh	Anupam Kher
Lata Mangeshkar	Smriti Mandhana	Shilpa Shetty	Lata Mangeshkar
Neha Kakkar	Harman Preet Kaur	Smriti Irani	Neha Kakkar
Dilip Kumar	Jemimiah Rodriguez	Shashi Tharoor	Dilip Kumar
Sachin Tendulkar	Mithali Raj	Farhan Akhtar	Sachin Tendulkar
MS Dhoni	Saina Nehwal	Sania Mirza	MS Dhoni

Calculating R_i for the set

Referring back to the formulae, looking at the stage 1, 2 and 3 proportion of occurrence, we calculated Ri for the set of names we were left with.

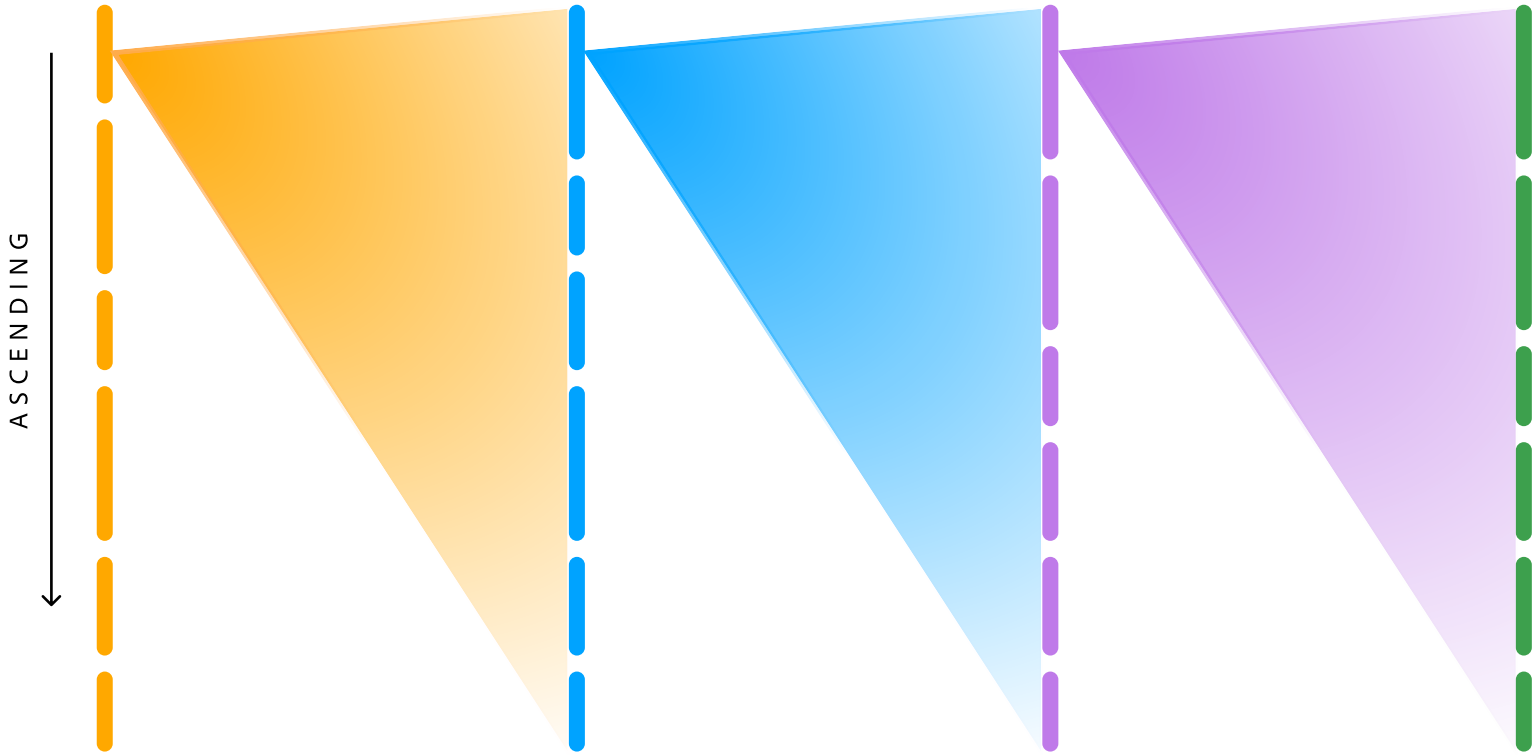
$$R_i = M_1*0.5 + M_2*0.3 + M_3*0.3$$

Where 0<Ri<1

Name	% Stage 1	% Stage 2	% Stage 3	Ri
A.B Diviliers	0	0.05	0	0.05
A.R. Rahman	0.04166666667	0	0	0.04166666667
Aamir Khan	0.125	0.05	0.3	0.475
Abhishek Bachchan	0	0.125	0	0.125
Adam Gilchrist	0	0.025	0	0.025
Adani	0.04166666667	0	0	0.04166666667
Aditya Thackrey	0	0.025	0	0.025
Adityanath Yogi	0	0.025	0	0.025
Aishwarya Rai	0.04166666667	0.1	0.3	0.4416666667
Ajay Devgn	0.08333333333	0.15	0.225	0.4583333333
Ajay Jadeja	0	0.025	0	0.025
Ajay Kapoor	0	0.05	0	0.05
Ajey Nagar	0.04166666667	0	0	0.04166666667
Akhhilesh Yadav	0	0.025	0	0.025
Akhilesh Yadav	0.04166666667	0.15	0.225	0.4166666667
Akshay Khanna	0	0	0.3	0.3
Akshay Kumar	0.2916666667	0.05	0.3	0.6416666667
Alia Bhatt	0.2083333333	0.05	0.3	0.5583333333
Alok Mishra	0	0.175	0.175	0.35
Alok Nath	0	0.025	0	0.025
Amir Khan	0.04166666667	0	0	0.04166666667
Amisha Patel	0	0.025	0	0.025

Ranking celebrities according to their R_i

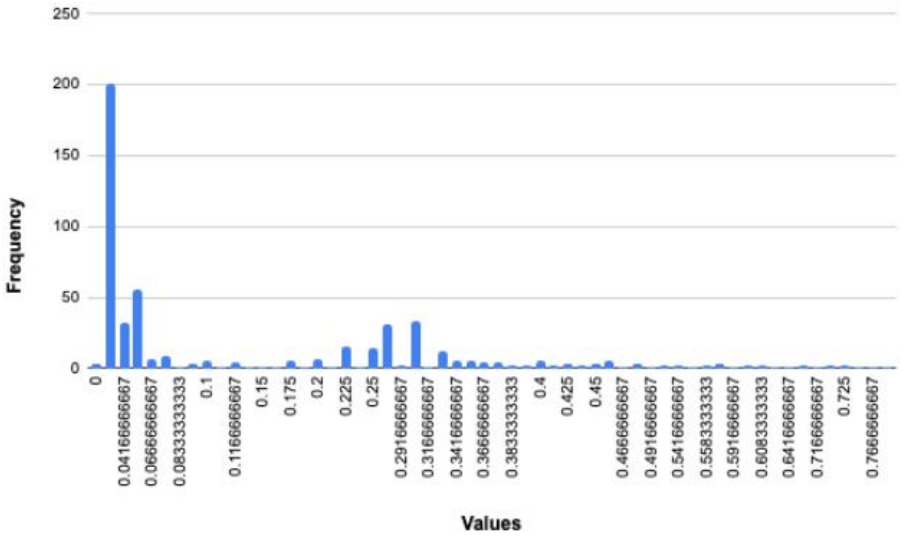
BREAKS REPRESENT UNIQUE VALUES



Creating Grids

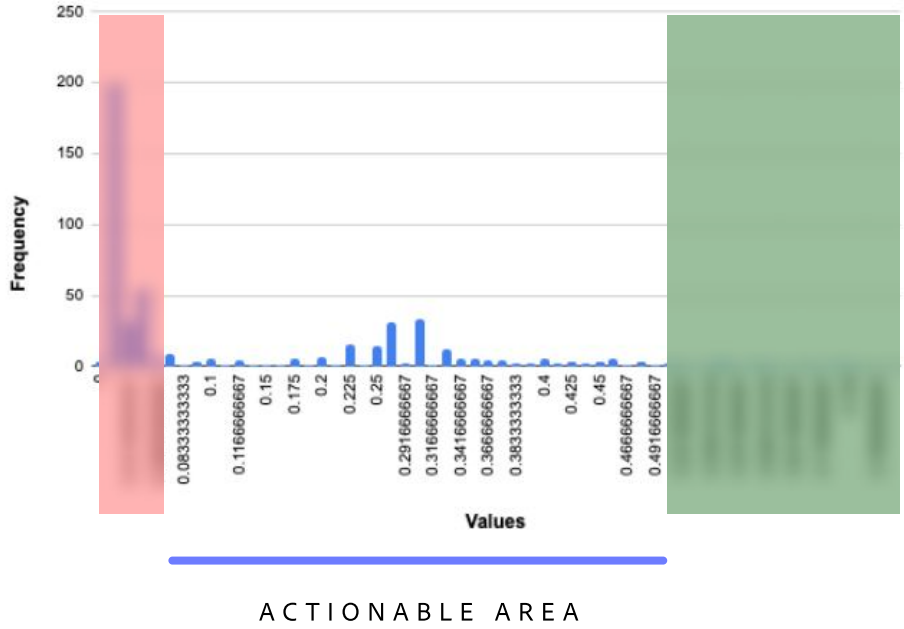
Distribution of Ri

After being left with a ranked list of names according to Ri, we looked at the distribution of frequencies across all the unique values. As hypothesised, the distribution was found skewed to the left, due to the diversity of names and less occurrence. There was a slight bulge in the middle and decreasing frequencies on the right end of the distribution.



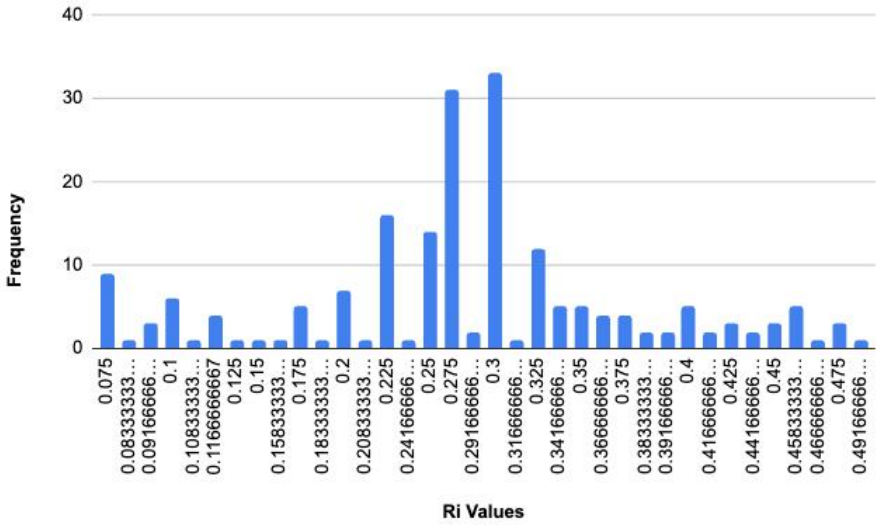
Trimming the list

As mentioned in the method earlier, we shall now trim the bottom 50 % and top 5 %. We don't want the entries with same values to be partially eliminated, hence we try to find which value is the is the closest to 50%ile and the 95%ile. These values came out to be 0.075 and 0.5 respectively



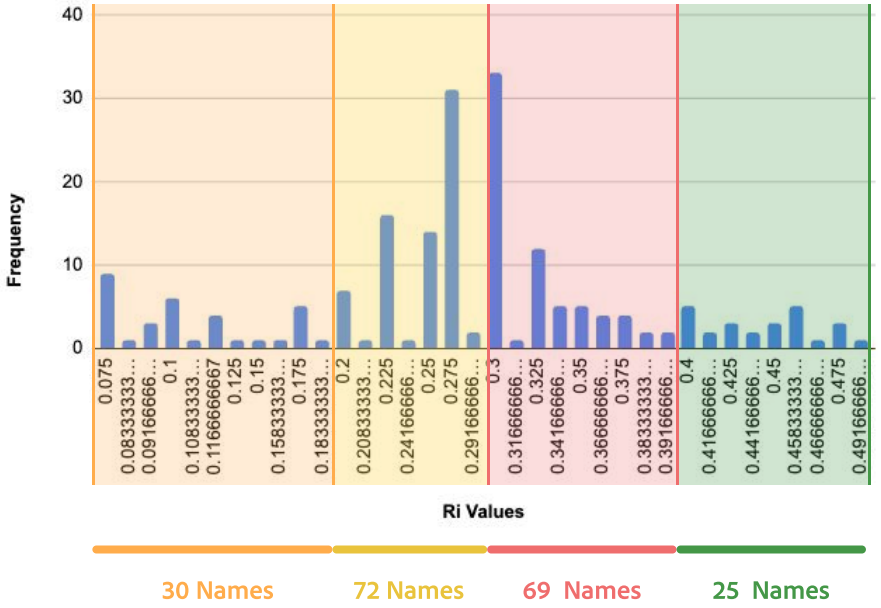
We were able to eliminate 299 names from the bottom and 27 names from the top. We were now left with 196 unique names from which we will create 4 grids.

Creation of Balanced Grids



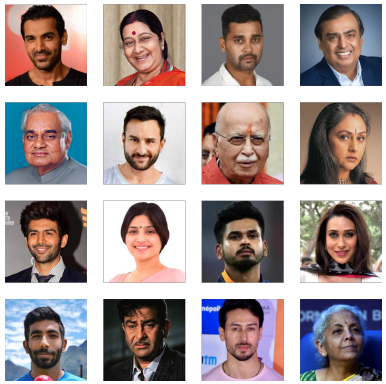
Left with 196 names now, we need to break down the list in 4 groups according to their Ri. The list was broken into following tiers:

- Tier 1: 0.075 < Ri < 0.2
- Tier 2: 0.2 < Ri < 0.3
- Tier 3: 0.3 < Ri < 0.4
- Tier 4: 0.4 < Ri



Now we had to construct 4 grids out of these 4 tiers. For this we randomly sampled 16 (Since it is a 4x4 grid) from each. The purpose of using random sampling instead of any logic was to keep the selection un-controlled. This makes this method reproducible for later studies.

Balanced Grids created using this method



Grid 1

Grid 2

Grid 3

Grid 4

Hence 4 grids were created using the method. These grids are regionally balanced which have celebrity with extreme or relatively low popularity removed and some local celebrities exclusive to the region of Kanpur. Hypothesis is that these grids will have a relatively random distribution of selection compared to grids used in previous studies, hence a higher entropy.

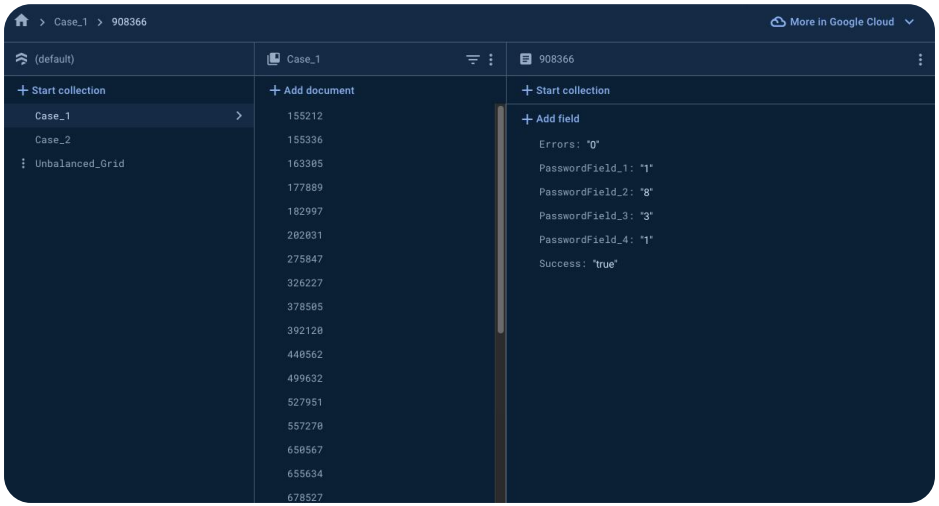
For evaluation of these grids we will compare them with grids used in previous studies (Zuha's study on passfaces). We will look at the frequency of each celebrity getting chosen and unique celebrities chosen by users.

UNBALANCED GRIDS

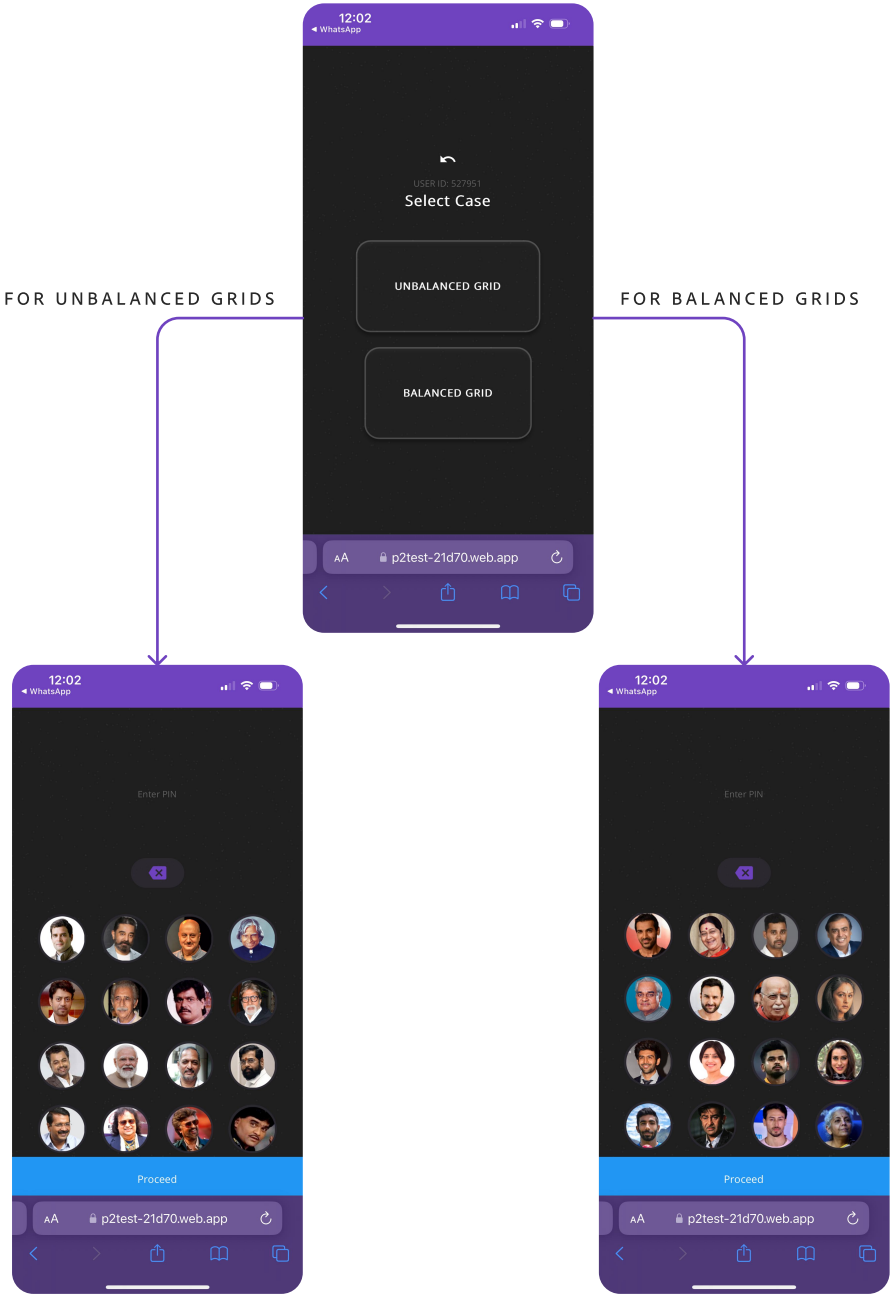
BALANCED GRIDS

Evaluation of grids

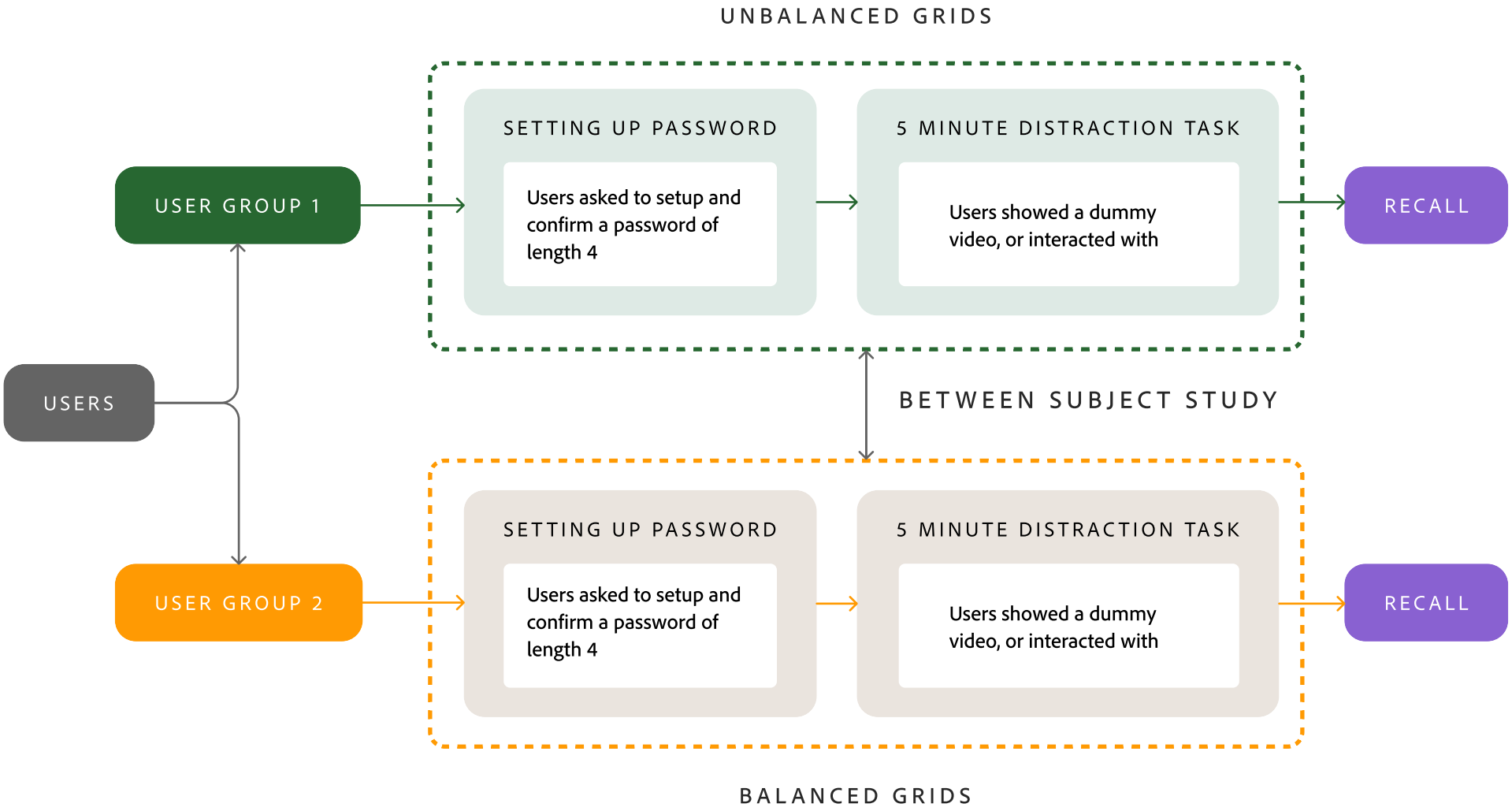
For evaluation, a web-app was created which had both the balanced and unbalanced grid. To preserve user's identity (as we don't need it for our study either), every user was assigned a user id. The responses were collected in Firebase realtime database and were exported to csv for analysis. The data collection for the study was also done in Kanpur. For this study, we only focus on short term recall, as the selection of celebrity is more relevant to the study. We also track error of input and success of recall.



SNAPSHOT OF DATABASE



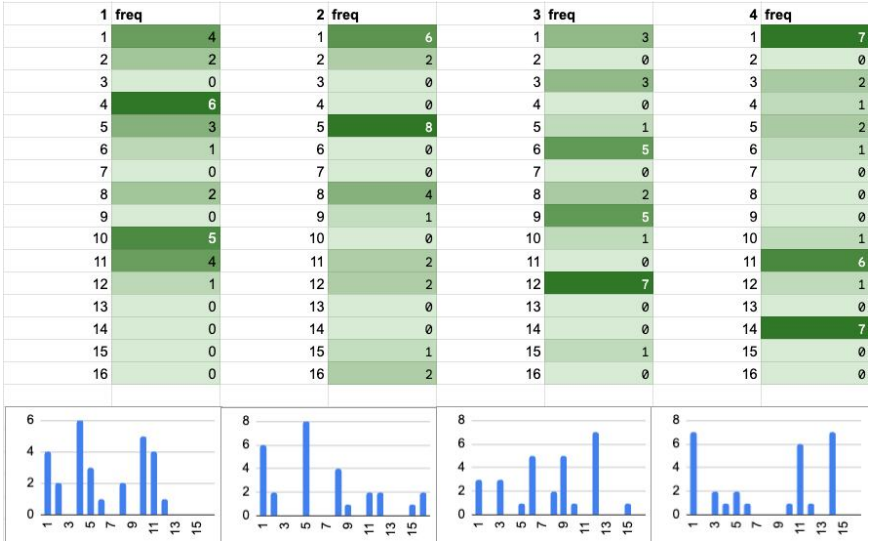
Experiment Design



Results after data collection

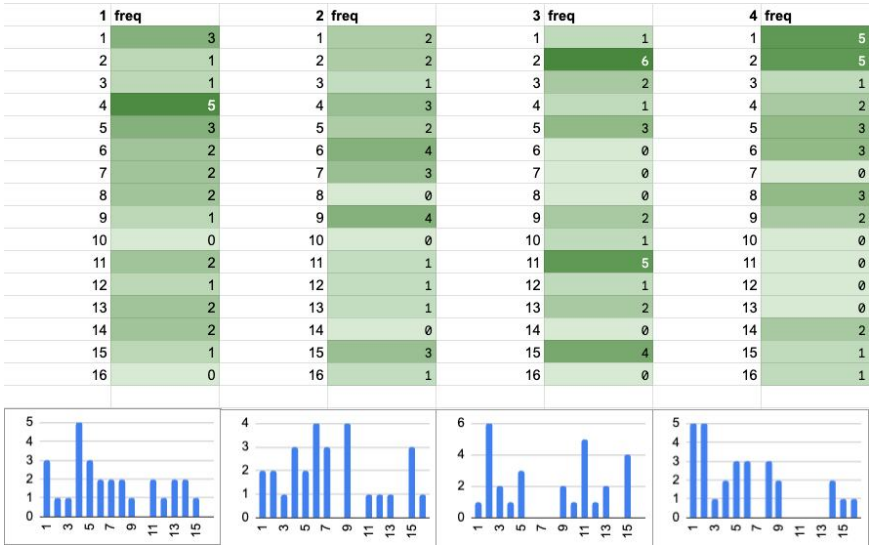
28 users were onboarded for each user group, 56 in total. The study was kept between the subjects to account for learnability. Apart from the faces selected, we also looked at errors in input and success in recall, for which the results were 100% for all the participants. (0 errors and 100% success rate). This can be attributed to the experiment expecting only short term recall. In figures given, we can see frequencies for the each index of the grid, along with distribution graph for each grid.

FOR **UNBALANCED** GRIDS ACROSS GRIDS



1	freq	2	freq	3	freq	4	freq
1	4	1	6	1	3	1	7
2	2	2	2	2	0	2	0
3	0	3	0	3	3	3	2
4	6	4	0	4	0	4	1
5	3	5	8	5	1	5	2
6	1	6	0	6	5	6	1
7	0	7	0	7	0	7	0
8	2	8	4	8	2	8	0
9	0	9	1	9	5	9	0
10	5	10	0	10	1	10	1
11	4	11	2	11	0	11	6
12	1	12	2	12	7	12	1
13	0	13	0	13	0	13	0
14	0	14	0	14	0	14	7
15	0	15	1	15	1	15	0
16	0	16	2	16	0	16	0

FOR **BALANCED** GRIDS ACROSS GRIDS



1	freq	2	freq	3	freq	4	freq
1	3	1	2	1	1	1	5
2	1	2	2	2	6	2	5
3	1	3	1	3	2	3	1
4	5	4	3	4	1	4	2
5	3	5	2	5	3	5	3
6	2	6	4	6	0	6	3
7	2	7	3	7	0	7	0
8	2	8	0	8	0	8	3
9	1	9	4	9	2	9	2
10	0	10	0	10	1	10	0
11	2	11	1	11	5	11	0
12	1	12	1	12	1	12	0
13	2	13	1	13	2	13	0
14	2	14	0	14	0	14	2
15	1	15	3	15	4	15	1
16	0	16	1	16	0	16	1

Selection Heatmap

For the grids used in previous study (Benchmark)

Darker color represents a celebrity getting selected more often

Grid 1

Grid 2

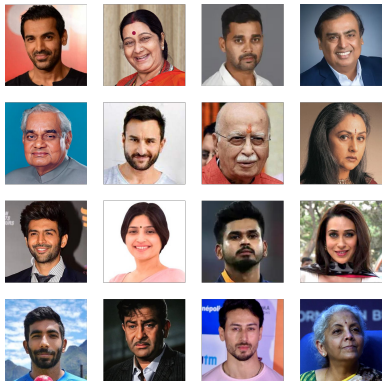
Grid 3

Grid 4

Selection Heatmap

For the **balanced grids** created using our method

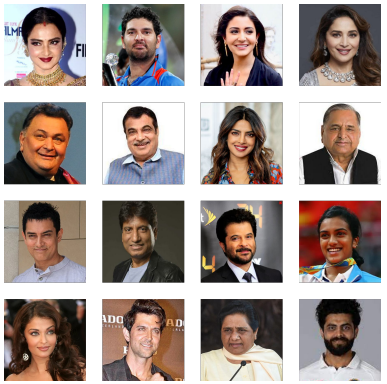
Darker color represents a celebrity getting selected more often



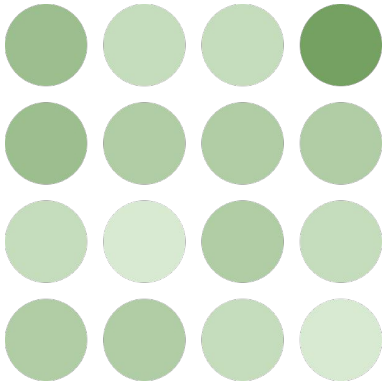
Grid 1

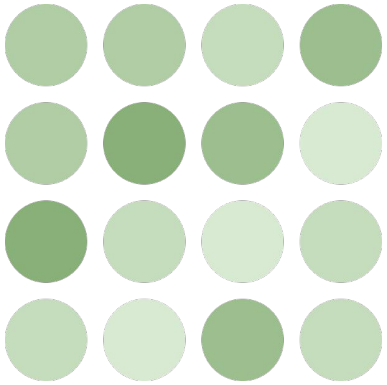
Grid 2

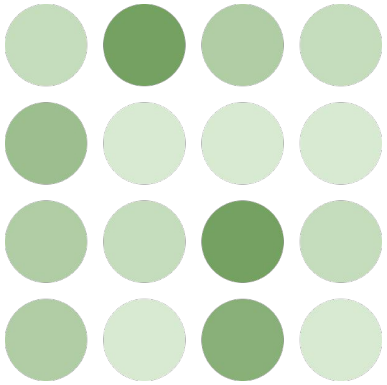
Grid 3

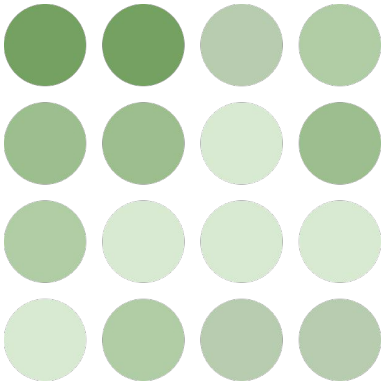


Grid 4





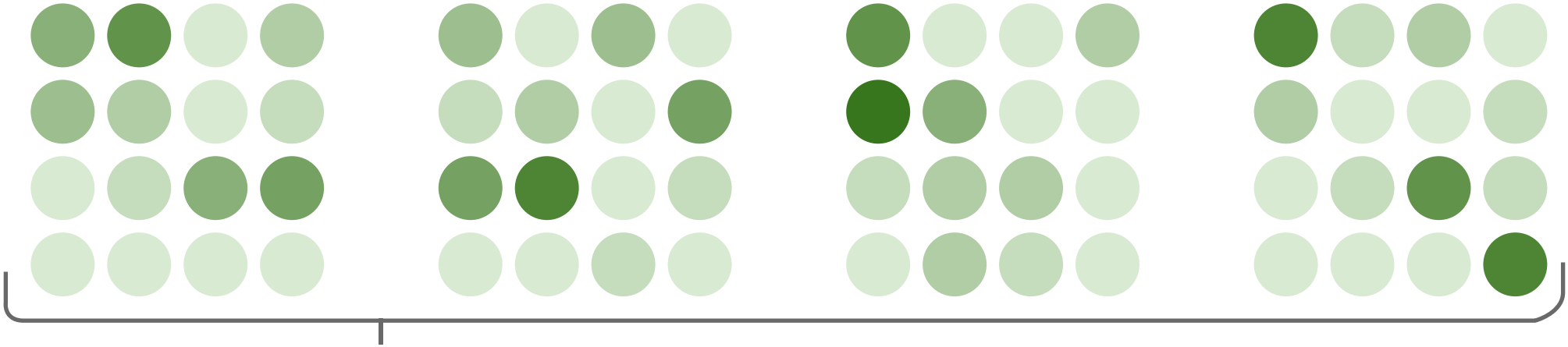




Selection Heatmap

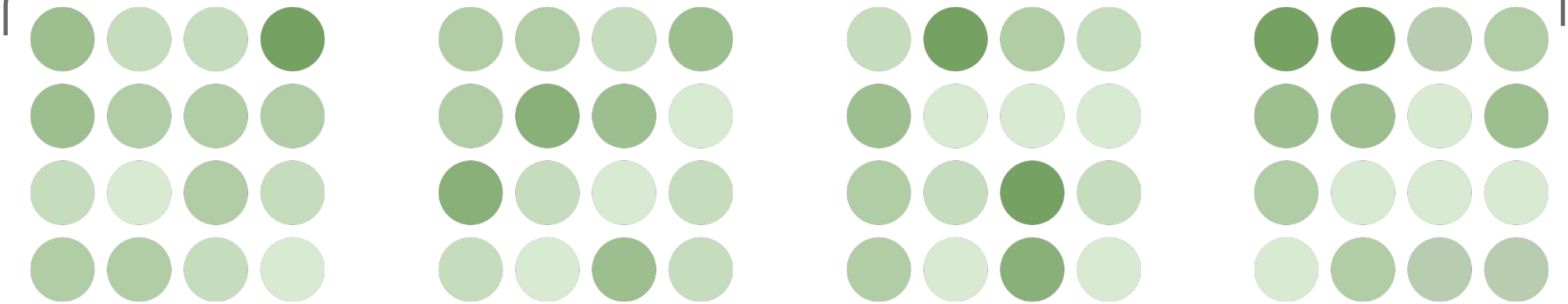
Visual Comparision

Darker color represents a celebrity getting selected more often



UNBALANCED GRIDS

BALANCED GRIDS



Comparing standard deviations

Standard deviation is a measure of the amount of variation or dispersion in a set of values. It quantifies how much the values in a data set deviate from the mean of the set.

A high standard deviation in a category indicates that the frequencies in that category vary widely, denotes lack of consistency.

As we can see, for unbalanced grids we get higher values of standard deviation for all grids when compared to balanced grids. **This suggests there is a uniformity in distribution of frequencies for each element in balanced grids created through our method when compared to the other set.**

FOR **UNBALANCED** GRIDS ACROSS GRIDS

1	freq	2	freq	3	freq	4	freq
1	4	1	6	1	3	1	7
2	2	2	2	2	0	2	0
3	0	3	0	3	3	3	2
4	6	4	0	4	0	4	1
5	3	5	8	5	1	5	2
6	1	6	0	6	5	6	1
7	0	7	0	7	0	7	0
8	2	8	4	8	2	8	0
9	0	9	1	9	5	9	0
10	5	10	0	10	1	10	1
11	4	11	2	11	0	11	6
12	1	12	2	12	7	12	1
13	0	13	0	13	0	13	0
14	0	14	0	14	0	14	7
15	0	15	1	15	1	15	0
16	0	16	2	16	0	16	0
Standard Deviation	2.049390153		2.380476143		2.236067977		2.54296415

FOR **BALANCED** GRIDS ACROSS GRIDS

1	freq	2	freq	3	freq	4	freq
1	3	1	2	1	1	1	5
2	1	2	2	2	6	2	5
3	1	3	1	3	2	3	1
4	5	4	3	4	1	4	2
5	3	5	2	5	3	5	3
6	2	6	4	6	0	6	3
7	2	7	3	7	0	7	0
8	2	8	0	8	0	8	3
9	1	9	4	9	2	9	2
10	0	10	0	10	1	10	0
11	2	11	1	11	5	11	0
12	1	12	1	12	1	12	0
13	2	13	1	13	2	13	0
14	2	14	0	14	0	14	2
15	1	15	3	15	4	15	1
16	0	16	1	16	0	16	1
Standard Deviation	1.238278375		1.341640786		1.879716291		1.693123347

Calculating Shannon’s Entropy

Shannon’s Entropy, is a fundamental measure in Information Theory. It quantifies the average level of 'information', 'surprise', or 'uncertainty' inherent in a message's possible outcomes. Entropy is calculated using probabilities of various events occurring, following the formula:

$$H(X) = -\sum_{i=1}^{n} P(x_i) * log_b P(x_i)$$

Where:

- H(X) is the entropy of the random variable X
- $P(x_i)$ is the probability of event x_i
- n is the number of possible events
- b is the base of the logarithm

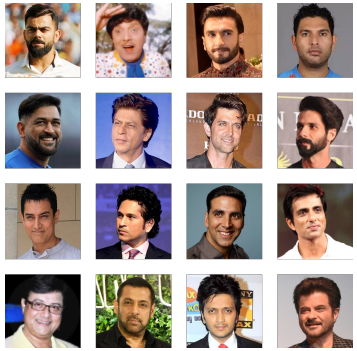
Higher entropy implies more randomness and less predictability and vice-versa

Shannon's entropy was calculated for both the cases for each grid. Higher entropy suggests more randomness in selection, which is our objective. It can be clearly seen that the values are higher for the balanced grids which were created using the method.

UNBALANCED GRIDS

Shannon's Entropy: ***0.89***

Shannon's Entropy: ***0.85***



Shannon's Entropy: ***0.86***

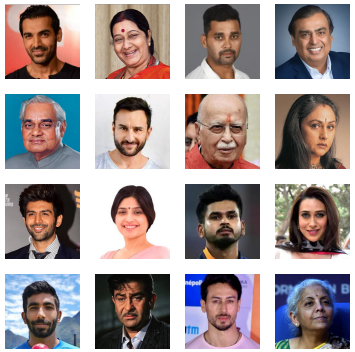
Shannon's Entropy: ***0.81***

Minimum reading
0.81

Maximum reading
0.89

Mean reading
0.8525

BALANCED GRIDS



Shannon's Entropy: ***1.05***

Shannon's Entropy: ***1.09***

Shannon's Entropy: ***0.95***

Shannon's Entropy: ***0.98***

Minimum reading
0.95

Maximum reading
1.09

Mean reading
1.0175

Applied paired Student's T test to the values

Dataset 1	Dataset 2			Dataset 1	Dataset 2
0.89	1.05		N	4	4
0.81	1.09		SD of the sample s	0.0	0.1
0.85	0.98		Observed mean	0.8	1.0
0.84	0.95				
			Paired t test	-4.464	
			Degrees of freedom	3	
			P value (two tailed)	0.0209	

H0: There is no significant difference in the set of values.

H1: There is a significant difference in the set of values.

p-value: 0.0209

Since, p-value<0.05

Null Hypothesis is **rejected.**

Hence,
There is a significant difference in the set of values.

Future Evaluation

At present, I am exploring a variety of additional evaluation methods. The investigation into their relevance and applicability is ongoing. Should these evaluations prove to be suitable, they will be incorporated into the final submission report. This ongoing research is aimed at enhancing the robustness and comprehensiveness of the study's findings.

Limitations and Future Scope

The method employed has its own set of limitations and drawbacks.

- **Robustness in Celebrity Listing and Ranking:** This aspect could be strengthened by increasing the number of participants, ensuring a more diverse range of names and a stable Recognition Index (Ri) for these names.

- **Systematic Approach to Grid Formation:** Though random sampling was used to create grids from the four tiers established post-trimming and ranking, this process could have been more systematic instead of being largely uncontrolled.

- **Arbitrary Decision-Making:** Certain cutoffs and values were arbitrarily decided, albeit with the understanding that they would be tested and updated until they reached a stable point. Nonetheless, a starting point was necessary.

- **Focus on Short-Term Recall:** The study primarily assessed short-term recall due to project contextualization and timeline constraints, resulting in 100% success rates. However, investigating long-term recall could reveal significant trends, especially when comparing balanced and unbalanced grids.

- **Randomisation of Layouts:** This aspect was not tested, as it wasn't part of the preceding study (Zuha's Study) either. Including layout randomisation could add another variable to the study, although it was omitted in this instance.

- **Grid Properties and Sample Size:** Unlike Zuha's study, which had grids with common properties (e.g., grids of middle-aged male or female celebrities), this study did not replicate that due to a smaller sample size of grids. Expanding data collection could mitigate this issue.

- **Inclusion of Regionally Popular Celebrities:** Despite the expectation that many celebrities popular exclusively in Kanpur would feature in the grids, only a handful of names consistently emerged. This suggests a need to explore systematic methods for including such names in the grid creation process, given that many were omitted due to the randomization process in grid making.

Conclusion

The employed method could produce balanced grids as we defined it. It was statistically proven that the grids created using this method had more randomisation in selection and a distributed frequency across the grid, when compared to unbalanced grids.

This study is easy to recreate but difficult to reproduce, which was something we aimed for. There was also inclusion of celebrities well known specifically to Kanpur. Hence it makes the grids regionally balanced.

Positioning of the Project

Within the domain of graphical pin-based password systems, this research endeavours to address prevalent gaps. The primary objective centers on the refinement of data collection methodologies, ensuring they encompass both recognition and recall aspects related to celebrity faces.

Additionally, a significant facet of this project seeks to provide a rigorous methodology for researchers which provides a criteria for evaluating the balance of celebrity grids and offers structured guidelines for their creation. By establishing these parameters, the research aims to contribute foundational insights and tools to enhance the rigor and effectiveness of graphical pin-based password systems in the broader academic and application contexts.

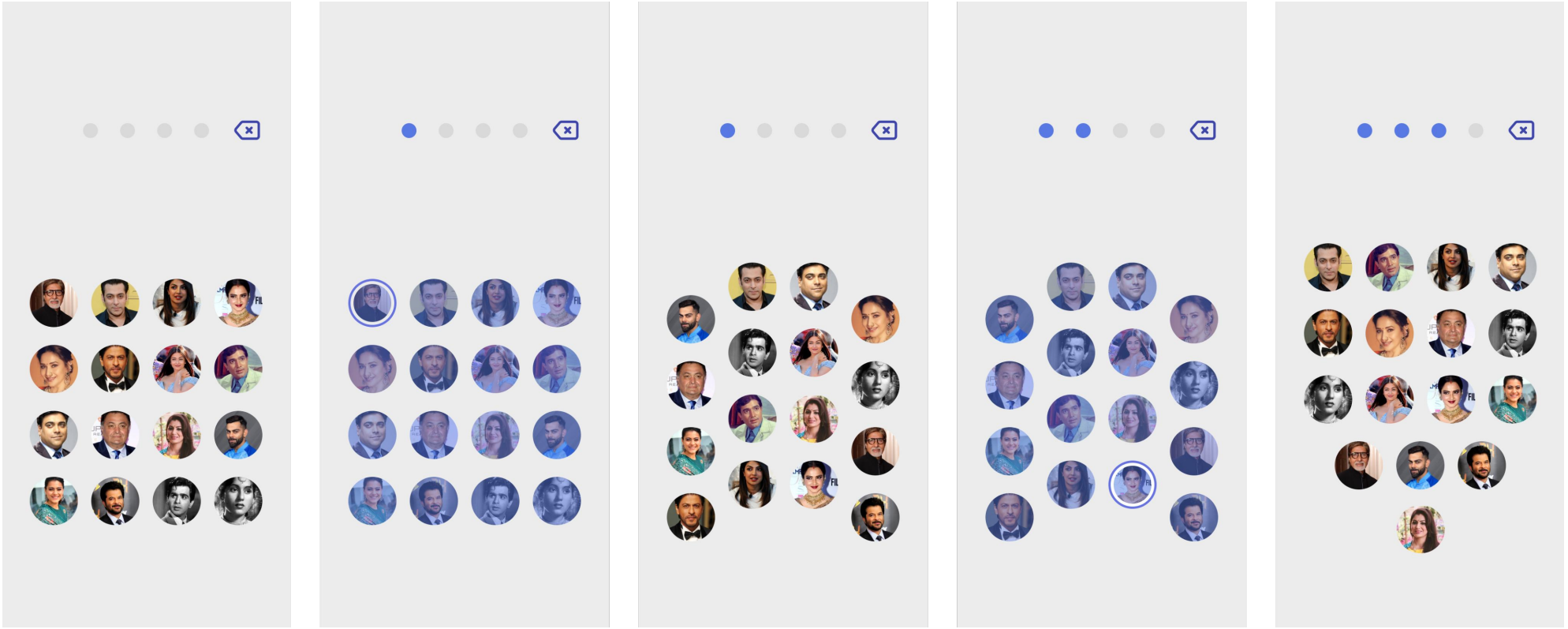
Motivations

My primary motivation for undertaking this project stemmed from a desire to deepen my understanding of quantitative data research methods. With a long-held aspiration to make an academic contribution, this project presented itself as an ideal opportunity for me to do so.

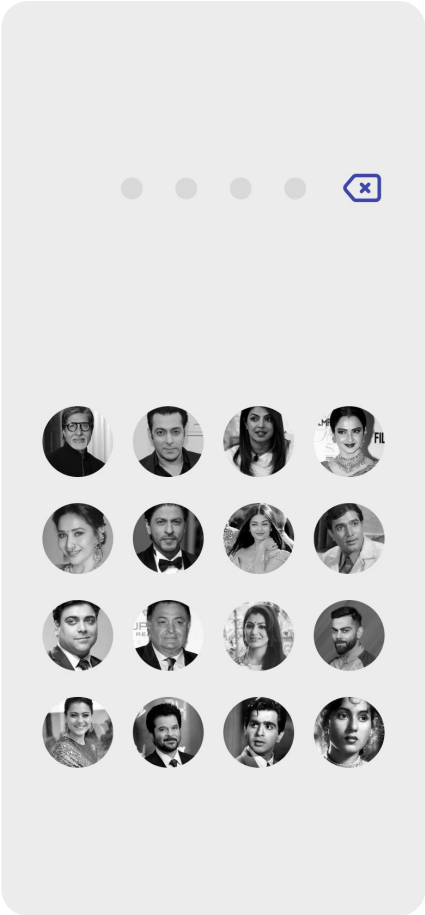
Given my background in computer science engineering, the realm of cybersecurity has always held a particular fascination for me. I view this project as a personal endeavor to attain proficiency and expertise in this discipline, leveraging my academic and technical background to explore and contribute to the field of cybersecurity.

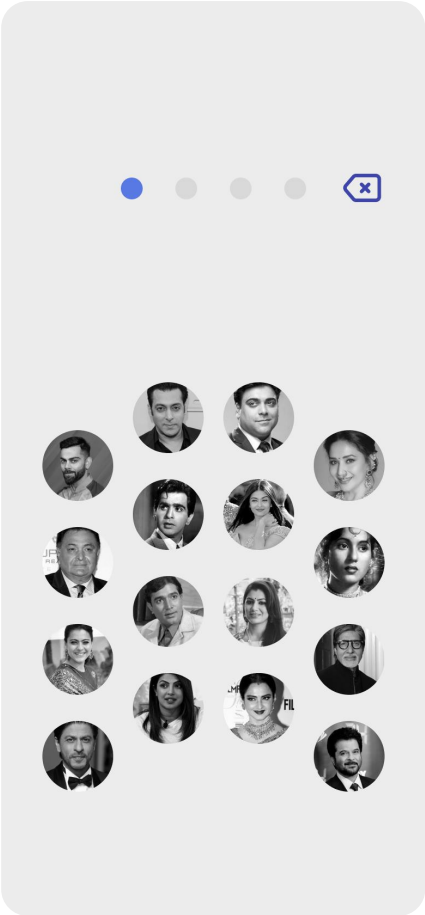
Study suggestions for shoulder surfing

1. Using Random Layout and immediate obscure after every input

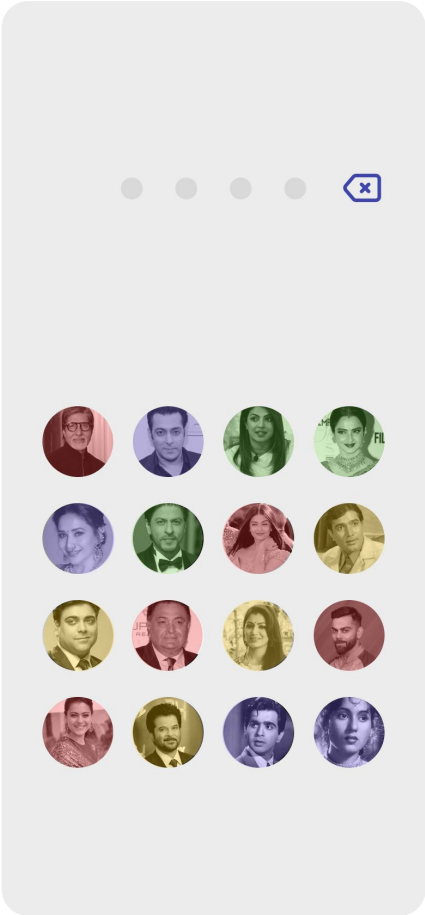


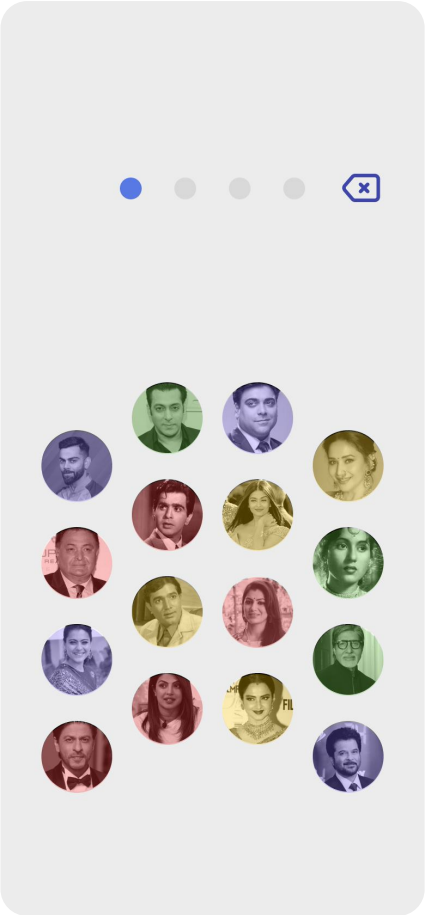
2. Using b/w filter to remove color aid for visual interpretation at a glance





3. Using random color filter for difficult visual interpretation at a glance





References

Graphical passwords: Learning from first 12 years - Robert Biddle, Sonia Chiasson, P.C Van Oorschot

Graphical password authentication using Pass faces Ms Grinal Tuscano, Aakriti Tulasyan, Akshata Shetty, Malvina Rumao, Aishwarya Shetty

Authentication Scheme for Passwords using Color and Text. B.O Vikas (2015)

Herley, C., van Oorschot, P., Patrick, A.: Passwords: **If we're so smart, why are we still using them?** In: Dingledine, R., Golle, P. (eds.) Financial Cryptography and Data Security, FC 2009. Lecture Notes in Computer Science, vol. 5628, pp. 230–237. Springer, Berlin (2009)

Anonymous Author(s). 2018. **Graphical Passwords for Emergent Users: A Four-day Recall Comparative Study on PIN**, Passfaces and Celebrities . 1, 1 (July 2018), 16 pages. https://doi.org/XXXXXXX.XXXXXXX // need to update this, not published yet

Raza, Mudassar & Iqbal, Muhammad & Sharif, Muhammad & Haider, Waqas. (2012). **A Survey of Password Attacks and Comparative Analysis on Methods for Secure Authentication.** World Applied Sciences Journal. 19. 439-444. 10.5829/idosi.wasj.2012.19.04.1837.

Melicher, William & Mazurek, Michelle & Kurilova, Darya & Segreti, Sean & Kalvani, Pranshu & Shay, Richard & Ur, Blase & Bauer, Lujo & Christin, Nicolas & Cranor, Lorrie. (2016). **Usability and Security of Text Passwords on Mobile Devices**. 527-539. 10.1145/2858036.2858384.

Anwar, Mohd & Imran, Ashiq. (2015). **A Comparative Study of Graphical and Alphanumeric Passwords for Mobile Device Authentication**. CEUR Workshop Proceedings. 1353.

Suo, Xiaoyuan & Zhu, Ying & Owen, G.. (2005). **Graphical Passwords**: A Survey.. 463-472. 10.1109/CSAC.2005.27.

yy