Does Eyewitness Guess or Recognize? Bootstrapping Face and Object Identification Accuracy

Kaja Glomb
Institute of Applied Psychology, Faculty of Management and Social Communication, Jagiellonian University, Krakow, Poland
e-mail: kaja.glomb@doctoral.uj.edu.pl

The purpose of the study is to determine whether the eyewitness identification can be regarded as a reliable source of information in a police investigation. In light of the many cases of eyewitness misidentifications, it seems reasonable to determine not only what class of objects is more likely to be actually recognized, but also is the level of accuracy sufficient enough to be a solid base for an investigation or a court case. To answer the questions a two-step experiment was designed and performed. At the first stage of the study, 71 participants watched a short video clip, and a week later they were asked to identify persons and the objects that appeared in the film. The participants’ rate of face identification success was 55%, while in the case of objects it was only 28%. Bootstrap estimation was used to determine if those numbers differ from random, and as a result whether they should be considered as a result of an accidental hit. The analysis showed that in the case of objects identification the success rate is within the bounds of randomness, while face identification exceeds it. It can be concluded that unlike faces, objects are more likely guessed than recognized.

Acknowledgements: The author thanks Jan Talaga from the Institute of the Middle and Far East for his help in organizing research and recruiting subjects. Thanks are also due to dr Iza Romanowska from Barcelona Supercomputing Centre for all insightful comments on the statistical analysis.


Introduction

The ability to recognize complex visual stimuli, although used every day, has consequences that go beyond common experience. It is the core of one of the most frequently used forensic procedures admissible in court – a police lineup. First-hand information acquired in the course of eyewitness testimony continues to have a certain importance that judges or jurors commonly attaches to, even though in the most dramatic cases mistakes made by an eyewitness during a lineup may lead to the conviction of an innocent person [32]. Therefore, it is hardly surprising to try to answer the questions that all parties to legal proceedings should be interested in finding: is eyewitness identification evidence that can be trusted, and if so, which category of visual stimuli has a greater chance of being correctly identified?

Face and object recognition or identification can be defined as a cognitive process that uses, on the one hand, visual stimuli received and analyzed by the primary visual cortex and, on the other, data stored in a long-term memory [2]. Recognition “happens” when a working stimulus is compared with its internal representation and, as a result of this comparison, it is possible to determine whether it is known or seen for the first time. Therefore, identification is a two-way dynamic process in which the subject receives stimuli and at the same time extracts information about them. It is performed by the recognition memory. Traditionally, recognition memory is divided into two different subsystems: recollection and sense of familiarity [14, 34]. The distinction is based on two qualitatively different functions. The first one is responsible for remembering the stimulus in its absence, as well as for recovering previously encoded information about it when the critical signal appears. This component leads to a full, complete recognition of stimuli and a context of its encounter. In contrast, familiarity can be described as a feeling that the stimulus was previously seen. It fulfills its function without extracting all of the related data. A sense of familiarity can take various forms on a continuum from an unclear feeling or intuition, to a compellingly strong belief that a given stimulus has already been perceived by senses [9]. More than 30 years of research [34] have brought a lot of evidence about the subjective experiences as well as objective indicators – including neuropsychological, neuroimaging, and neurophysiological studies of humans and animals, that support this distinction evidenced by partially distinct neural substrates of those two components [13].

However, as Medina [20] points out, the two-component model can be replaced by a hypothesis of memory strength understood as a continuum from "sense of familiarity" to full "recollection". This approach appears to be close to the concept of memory formulated by Damasio [4, 22], who postulates abandoning the metaphor of “memory warehouse”, where all of the information is stored in favor of a concept of memory reconstruction. He believes that internal representations of real-life objects are not stored anywhere permanently, but rather temporarily replicated based on previously seen patterns. It happens when the critical signal activates the neural network corresponding to the sensory representation of the object. Thus, this concept assumes that cognitive representations are not a reliable representation of real objects, but rather an approximate picture of them. Damasio’s hypothesis seems to accurately explain the eyewitness misidentification – their internal representations of faces or objects are results of a single, often short episode of exposure to visual stimuli, which may not be sufficient to trigger extensive neuronal networks necessary for full recognition.

With this in mind, It can be hypothesized that when witnessing a crime event, a person may not be able to encode all visual information and full context of stimuli, thus during the lineup
identification, he or she tends to select an object the most resembling their internal representation. This strategy may work with some class of objects, but not necessarily all of them.

The analysis of the extensive body of knowledge of eyewitness testimony suggests that information about a crime is remembered differently depending on whether they concern central or peripheral details. During a crime event, eyewitness usually experience negative emotions that ultimately leads to the attention narrowing and a phenomenon of tunnel memory [e.g. 3, 8, 15, 29, 30]. As Reisberg and Heuer [24] argue, the effect may result from the occurrence of powerful “attention magnets” present within most of the emotional events. To be considered a central, the detail must: (a) capture the eyewitness attention; (b) constitute an integral part of the stimulus (spatially, temporally, conceptually); (c) be relevant to current goals [19]. The examples of magnets used in experiments vary from a single object, e.g. a bloody face or a weapon [16, 17], to a more complex scene [25, 26]. However, as Laney and her colleagues suggest [17], the real-life emotional memories don't necessarily have magnets that are the immanent, internal visual foci of the memory - the emotional weight of memory is rather the result of its “thematic” properties rather than visual (Laney et al., 2003). It is also likely, that eyewitness memory depends on the differential salience attributed to the stimulus. Some characteristics of the event may be forgotten not because of the attention narrowing caused by memory magnets, but as a result of the subjective impression of stimulus insignificance from the eyewitness's point of view.

It is also important to note that the vast majority of experiments on central v. peripheral details of the crime are interested in recollection not recognition. However, as Deffenbacher's [6] points out, memory performance differs depending on whether the eyewitness is asked to describe certain events or to identify people or items involved in it. Moreover, the research on eyewitness identification seems extensively concentrated on recognition of human faces, while occasionally, eyewitness is also asked to identify inanimate objects. Yet, our knowledge of this aspect of the memory seems to be insufficient, especially when considering the responsibility that rests on eyewitnesses and their testimonies, often determining the direction of the police investigation. The question of whether the accuracy of identification of inanimate objects, which are less important for the course of the event, therefore seems to be justified.

With this in mind, the experiment comparing the accuracy of recognition of central v. peripheral stimuli was performed. It was presumed that the faces of actors playing in the scene presented to the subjects are central stimuli, while peripheral details are items connected to the actors (i.e. clothing) or a background element. The rationale behind the assumption that human faces will be the attention foci, while inanimate objects may be considered peripheral is that people show a natural interest in other's faces. The phenomenon is not only evidenced by infants’ ability to recognize and follow faces of their caregivers [e.g. 23], but also our tendency to notice faces even in abstract material, i.e. best demonstrated by emoticons created from punctuation marks.

To determine what class of stimuli is more accurately recognized by eyewitness, thus can be regarded as a credible testimony, it was also necessary to define a condition that would constitute such confidence. In the case of this study, the minimum criterion is the threshold of the randomness of answers given by eyewitness asked to identify a culprit or an item. In other words, it can be assumed that the identification of witnesses may - although not necessarily - be correct, because in general, objects of a given class tend to be recognized rather than "guessed" or pointed randomly. Referring to the issues indicated earlier, the following research hypotheses were formulated:

H1. Both faces and object identification accuracy rate will be above the threshold of randomness identified as an indicator separating results that are most likely the effect of "guessing" from those
that were "identified". The hypothesis is supported by studies on the capacity of long-term memory, indicating that people have a high ability to store detailed information about faces [e.g. 5] as well as real-world objects [e.g. 1]. However, it is expected, that the level of accuracy of recognition will be lower when the experiment is conducted in the eyewitness identification paradigm, i.e. when time delay and low level of knowledge of the objects (seen only for a short period of time) is taken into account.

**H2.** There is a qualitative difference between the accuracy of facial and object recognition which reflects the differences in the stimuli processing and neural substrates of these processes [12]. Based on the previous studies of eyewitnesses’ testimonies, indicating that the memory of central issues is less susceptible to distortions, the hypothesis assumes that faces will be identified more accurately than objects.

To verify those hypotheses an experiment that would imitate natural memory processes and refer to the processes occurring in the memory of eyewitness was design and performed.

**Research Program**

Several assumptions which determined its final course of the research program was made: (1) faces and objects are to be exposed in the natural environment, not abstracted from the context and presented individually, (2) both faces and objects are to be presented in the same conditions and for the same period, (3) the subjects are unaware of the purpose of the experiment, i.e. memory testing, which will allow to study the natural processes of remembering and forgetting of visual stimuli. In order to meet these criteria, appropriate materials were prepared to reflect the legal context of the research.

**Procedure.** The study was designed as two stage experiment that to some extent imitates the experiences of a typical eyewitness. It began with the coding of visual stimuli presented in form of short video clip. After watching it, the participants answered a masking question ("What in your opinion could have happened shortly after the video ended?") – it was used to cover the actual purpose of the experiment, and avoid using any mnemotechnic that could facilitate further recognition. Week later, the subjects were asked to identify 6 stimuli from lineup photo array. Unlike real crime events, they were not requested to describe objects of identification – some studies indicate that primary face description can influence decisions made during the lineup, causing a verbal overshadowing effect [11]. The experiment was conducted in group condition.

**Eyewitness event.** At first stage of the experiment, participants watched a 3-minutes long video clip presenting a scene in café. They saw a subtle interaction between two "characters" - a woman and a man, who sit at adjacent tables. In the background between them hangs a decoration – a colorful Christmas sock. The woman and the man do not communicate verbally with each other, although at some point the man taking advantage of the fact that the woman is leaving the table for a moment, throws a mysterious message into her purse. Both, the man and the woman are exposed for the same amount of time (c. 2'30" +/- 3"). The actors were ordinary-looking, young adults in a similar age to participants and same race in order to avoid own-age [28] and same-race bias [21], with no distinctive features enabling easier identification [18, 31]. They were amateurs, unknown to

---

1 The study uses materials and data obtained during the experiment on the influence of the cognitive warm-up effect on face and object recognition conducted by the author. At the time of submitting this article, the results of that experiment have not yet been published. The dataset has been made available online at https://doi.org/10.6084/m9.figshare.9741398.v1
subjects prior the experiment. Objects later used in identification phase belonged to the actors or were in the café’s equipment.

**Lineup.** Six sets of photo arrays (app.), representing lineups, were prepared for the experiment: 1) male face 2) female face 3) baseball cap 4) male hoodie 5) female sweater 6) Christmas decoration. Every lineup was created of a realistic, high quality photographs of faces or objects that were previously seen in the video clip and 3 similar looking fillers. Then creating the photo array lineup current regulation of the polish Ministry of Justice on technical conditions during police lineup were applied. Actors as well as fillers were asked to take a picture in neutral pose, similar to typical mugshot. Pictures of objects comparable to those used in video clip were obtained via Google Reverse Image Search. The photographs were arranged in one horizontal row. In the case of face images, the 'correct face' were placed on second or third position, while items were placed on all four positions.

**Participants.** A total number of 71 participants took part in the study. They were recruited from among university students of cultural studies in Cracow, Poland. The vast majority of the participants were women (81.7%). The youngest were 19 years old while the oldest one 23. The mean age was 19.8. Almost all of them had secondary education with just one who held a bachelor's degree. All participants reported good health and had normal or corrected vision. They were aware that they could resign from participation in experiment at any stage.

**Results**

In this section, analyses are reported for two aspects of lineup performance: (a) the accuracy of face identifications, (b) the accuracy of objects identification. The table below shows the number and percent of participants who correctly identified the stimulus from a lineup.

<table>
<thead>
<tr>
<th>A number and percent of subjects who correctly identified stimuli from a lineup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>%</td>
</tr>
</tbody>
</table>

**The accuracy of face identifications.** In order to assess the accuracy of the face identification a ratio of correct answers in lineup performance to the total number of attempts was taken into account. This parameter, defined as the success rate, allows us to answer the question concerning the level of randomness of the obtained results. Considering the fact that 71 subjects were to identify two faces from two sets of the lineup made of 4 photographs, and they chose the correct photograph 79 times (42 hits in female face, 37 hits in male face), the success rate was 0.55 or 55%. At first glance, it seems that the success rate is far from random. In case of procedure that requires choosing one of four photographs, the success rate of .025 would indicate that the face was chosen randomly – “guessed” rather than “identified”. However, to be sure, bootstrap estimation was carried out. This type of analysis, increasingly popular in social sciences, has many advantages in studies involving small groups that bring data prone to having non-normal distributions. Bootstrapping as a method of calculating standard error or confidence interval, therefore estimating the precision of any statistics, requires fewer assumptions than more traditional approaches and
gives a stronger and more general conclusions [33]. Fundamentally, it assumes that the observed data are a representation of the underlying population characteristics, so new samples, each of the same size as the observed data are drawn with replacement from data observed. The final statistic is calculated for each new data set.

In this study, bootstrapping was used to estimate the 95% percentile confidence interval of success rate – 10 000 bootstrap samples were drawn, and the success rate was calculated in each of them using Python™ programming. The 95% percentile bootstrap confidence interval ranged from 0.17 to 0.32 (fig. 1). That is, if the participants chose the photograph at random, we would expect that the number of hits falls between 17%-32% in 95% of cases. In the experiment it was 55% thus the participants were about twice as likely to recognize the face than if they were guessing at random. The result is significant; therefore, it is legitimate to conclude that faces are actually recognized, not chosen randomly.

The accuracy of object identifications. A similar procedure was performed to determine the randomness of object identification. The total number of correct identifications was also 79 (17 hits in case of a cap, 15 hits in case of a hoodie, 24 hits in case of a sweater and 23 hits in case of a decoration) but the number of ‘trials’ was twice time bigger – the subjects had to recognize four objects from four different lineups. Consequently, the success rate was 0.28. Again, 10 000 bootstrap samples were drawn. In the case of object identification, the 95% percentile bootstrap confidence interval, ranged from 0.2 to 0.3 (fig 2.). That is, if the participants chose the picture at random, we would expect the number of hits to fall between 20%-30% in 95% of cases. In the real experiment it was 27% thus it is not likely that the number of hits differed from a random scenario. The result is not significant, therefore, the hypothesis that the subjects randomly selected photos during the identification test must be accepted.
Figure 2. Probability density function of the object identification success rate distribution

Referring to the hypotheses set out earlier, the results only partially support the first hypothesis. The success rate in case of facial identification is above the threshold of randomness, and at the same time, the experiments showed that objects are most likely randomly chosen from a photo array. However, the experiment confirms the hypothesis suggesting that the success rate is higher for central details.

Conclusion

The study objective was to compare the accuracy of central v. peripheral details recognition in the eyewitness identification paradigm. It was hypothesized that faces were central details, while inanimate objects were considered peripheral. The comparison allows to answer the question whether the identifications of faces and objects made by eyewitnesses can be trusted and regarded a valid evidence in legal proceedings. The minimum criterion for considering this type of testimony credible was a randomness threshold that indicates the probability with which the witness is guessing. The study shows that whether the identification by witnesses can be considered trustworthy depends on the class of the recognized object. In case of face identification, the success rate achieved by the participants exceeded the randomness level and meets the minimum criterion. Although the criterion defined this way can be considered arbitrary, it is difficult to imagine another, more impartial one – there is no objective justification for which the success rate should exceed e.g. 50, 60 or 70% to be considered reliable. Nevertheless, it is important to reflect on the level of ability to recognize unknown faces observed in the study. The result – 55% success rate – achieved by the subjects seems to be quite average. They did not know the purpose of the study and were unaware that in second stage of the experiment they are going to be asked to identify faces, thus during the lineup they used memory resources not influenced by some variables, that in the case of real witnesses increase the probability of misidentification. Although it is impossible to
compare any experiment with the experiences of actual witness – the burden of their decision is incomparably higher – the participants of this study found themselves in an unknown, surprising and intriguing situation. In some aspects it imitated the task faced by actual witness, so the results can be considered as a reflection of the memory performance.

The study also showed that the credibility of witnesses in identifying items is particularly low. The success rate achieved by the participants in the object identification test is within the bounds of randomness. It means that when a witness is asked to choose a photograph presenting an object they saw during an event under investigation, with high probability, he or she rather guesses, not necessarily recognize it. While interpreting the results, one should take into account the fact that the objects recognized by the subjects were insignificant for the events taking place during the video clip, thus it is possible that items attracting attention or evoking emotions may cause perceptual narrowing/tunnel vision, resulting in transferring of attention resources and better memorizing.

Taking into account the experiment design as well as the fact that the scene did not reflect the wide range of emotions experienced by a real eyewitness, an important question concerns the reason for the difference in accuracy of identification between central faces and peripheral items. The study suggests that holistic processing may be more effective for the purpose of recognition. It allows us to make a basic distinction between known and unknown objects, which is possible without prior analysis of particular features. Therefore, it is possible that holistic processing increases the chance of correct identification of the perpetrator during the lineup. At the same time analytical processing seems to be insufficient to distinguish between known and unknown objects. One can say, referring to Damasio’s theory, that the memories of faces are more likely to be correctly replicated possibly due to the fact that a critical signal activates extensive neural networks in case of objects processed holistically, which is not the case for analytical processing.

Although holistic processing is usually associated with faces, it is not limited only to them. Some research on the so-called expertise hypothesis suggests that as knowledge of the category of objects increases, it is more likely that they will be processed holistically [e.g. 7, 11]. Thus, it allows us to consider identification made by someone with substantial experience in particular objects recognition as credible – a car that a suspect drove off with is more likely to be correctly identified by a car dealer than by a casual passer-by. However, since law enforcement is probably rarely lucky enough to find those expert-witnesses, it is necessary to emphasize the need to approach this type of testimony with great caution. It may protect police and from tunneling their attention and investigatory efforts to the object misidentified during the lineup.

Although the study has shown that the people are better at recognizing faces, the accuracy of identification does not appear to be convincingly high. Therefore, it is worth considering whether in criminal cases, which have such great social significance and serious real-life consequences, face
identifications should be treated as valuable judicial evidence. Perhaps a discussion should be launched on how the relevance of such evidence should be assessed by the jury or court. The decision on whether a particular piece of evidence can be considered credible is taken in every single case and it is guaranteed by the principle of free evaluation of evidence. Thus, it seems essential to raise awareness of the limitations of human recognition memory among practitioners of law and those involved in judicial proceedings.

References
10. Galton F. Composite Portraits, Made by Combining Those of Many Different Persons Into a Single Resultant Figure // The Journal of the Anthropological Institute of Great Britain and Ireland. 1879. No. 8. 132-144 pp.
17. Laney C., Heuer F., & Reisberg D. Thematically- induced arousal in naturally occurring
32. Wells G. L. Eyewitness identification. 2018. Online version: https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1066&context=psychology_pubs

Information about the authors
Kaja Głomb, PhD student in Social Sciences/Psychology, Institute of Applied Psychology, Faculty of Management and Social Communication, Jagiellonian University, Krakow, Poland, e-mail: kaja.glomb@doctoral.uj.edu.pl

Получена 03.06.2020
Принята в печать 07.09.202_

Received 03.06.2020
Accepted 07.09.2020
Appendices