Running head: Relative judgment. When the relative judgment theory proved to be false

Levi A.M., PhD, On pension from Israeli Police (avmlevi@netvision.net.il)

A commonly accepted theory is that when witnesses can identify culprits in lineups, they will concentrate on him. On the other hand, when they cannot they compare between lineup members and choose the person most similar to the culprit. Therefore they will divide their gaze more equally between foils. An eye tracker was used with a 48-person lineup (four screens with twelve photos in each) in an attempt to demonstrate the superiority of gaze behavior over the verbal response. Surprisingly witnesses usually concentrated on some foil as much as they did on the target. Alternate theories are required to explain the reduction of false identifications in sequential lineups. The advantage of large lineups was demonstrated. Police may use them in conjunction with eye trackers to find culprits that witnesses focus on despite saying that they are absent, the only known method to increase correct identifications.

Keywords: relative judgment, eye tracker, large lineup, identification, gaze behavior, simultaneous lineup.

      The lineup is a procedure in which a person suspected by the police of having committed a crime is shown the suspect, along with a number of known innocent people ("foils"). If a witness chooses the suspect, this is taken as evidence of his guilt by the courts. The lineup is the safest eyewitness identification procedure. However, it is far from perfect. There is ample evidence that witnesses often choose someone who is not the culprit (Conners et al., 1996; Scheck, Neufeld, & Dwyer, 2001, Wells et al., 1998). When they choose someone who is not the suspect but a known innocent, the police know that they have erred. However, by chance witnesses choose a suspect who is innocent 1/N times, where N is the lineup size. With the common American lineup size of six, this will happen 1/6=0.167, or almost 17% of the time.

      There is a second error that witnesses often make which goes undetected by the police: witnesses fail to identify guilty suspects (Levi, 1998). While a number of innovative lineup procedures have been developed to reduce mistaken identifications (Levi, 2006a; Levi, 2012;

Lindsay & Wells, 1985; Pryke et al., 2004), there have been no procedures available to increase correct ones.

The eye tracker is a device that photographs the movement and location of the eyes' gaze at some stimulus (Holmquist et al., 2011). It was used by Loftus, Loftus and Messo (1987) in an eyewitness study, and more recently by Brace (2011), Pike (2011), Hunter and Pike (2011), and Mansour et al. (2009).

The research hypothesis of Hunter and Pike (2011) was that the gaze of witnesses looking at the culprit in a lineup which resulted in an accurate identification would be different from the gaze of witnesses who chose someone else, potentially an innocent suspect. If this were true we might be able to dispense with the unreliable verbal response of the witness and base identification decisions on their gaze pattern instead, increasing correct identifications and decreasing mistaken ones. The results were encouraging.

Hunter and Pike (2011) used the relatively new English lineup procedure, which is unique. English law forbids conducting photo lineups, where photos of the lineup members replace the actual members, yet conducting live lineups is a very difficult procedure entailing a great waste of resources. The English solution has been, instead of taking photographs of suspects' faces, to take short video clips of them, where they move their head slowly from side to side. An appropriate sample of these video clips from past cases is chosen to be the foils in the lineup. Along with the present suspect, they are shown sequentially to the witness, at least twice.

While no clear theory seemed to predict gaze behavior for the English lineup, there seemed to be interesting possibilities for the traditional simultaneous lineup. In the simultaneous lineup witnesses view the entire lineup with all its members seen at the same time.

According to a popular conceptualization espoused by Wells (1984) (relative judgment), witnesses with poorer memory of the culprit compare between lineup members, and often simply choose the person who seems to look most like the culprit-often the innocent suspect. Translating this into gaze behavior, in comparing between lineup members the attention of these witnesses will be on some of them in addition to the time spent concentrating on the culprit. Perhaps more time will be spent gazing on the culprit, but not a tremendous more time.

On the other hand, witnesses with relatively good memory of the culprit are expected to spend far less time gazing at the other lineup members. Indeed, witnesses using this "absolute" strategy tend to spend less time in making their identification (Sporer, 1993). This less time should be concentrated far more on the culprit. Translating this into gaze behavior, while they might be expected to at least glance at the other lineup members, a reasonable expectation would be that they spend much more time looking at the culprit than at any other lineup member.

The more important different behavior occurs when the culprit is not in the lineup, when the suspect is innocent. Witnesses with good memory should be able, after glancing at the lineup members, to decide that the culprit is absent. Other witnesses will compare between lineup members and choose the person most resembling their memory of the culprit, who all too often will be the innocent suspect (unless of course all the lineup members differ tremendously from the culprit, which should not happen in a fair lineup).

These conflicting predictions lead to a promising outcome: witnesses who dwell a relatively long time on the suspect have identified the culprit. On the other hand, if the suspect who is chosen does not stand out as having been looked at so much longer than any other lineup member, he/she was most likely chosen using relative judgment, and therefore is likely innocent.

This analysis differs from that of Mansour et al. (2009). That paper states that if a witness looks at all the faces in a lineup, this is indicative of relative judgment. This position contrasts with this paper, which expects witnesses to at least glance at all the faces. Relative judgment is indicated only if the witness fails to focus much longer on the person chosen.

Levi (2012) has introduced the large lineup. In a series of experiments he compared lineups of 42, 84 and even 120 with lineups of about 20, and found no difference in either correct identifications or the number of witnesses who mistakenly chose someone in culprit-absent lineups. A vital common element in these large lineups was that no more than twelve photos were displayed on each page. Thus, the 84-person lineup consisted of seven pages, the 120-person lineup of ten pages. A lineup of 168, with 24 photos of each of seven pages, resulted in markedly fewer identifications, as did a pilot study of 18 photos per page.

The clear advantage of large lineups is that they reduce the chance of mistaken identifications. Steblay et al. (2001) found that sequential lineups reduced mistaken choices to an average of 28%, compared to the 51% for simultaneous lineups. However, in the standard American lineup of 6, this means that 28/6 = 4.7% innocent suspects are mistakenly identified. This compares to a relatively small large lineup of 48 with an average of 51% mistaken choices, leading to 48/51 = 0.9% innocent suspects mistakenly identified, 5 times less than for the sequential lineup. The number of mistaken identifications in the six-person simultaneous lineup is 51/6 = 8.5, almost ten times more than for the 48-person lineup.

This experiment used indeed a 48-person lineup, rather than a 6-person simultaneous one. A major reason is to begin gathering data on the much superior large lineup. Another is to provide a tougher test for the eye tracker, forcing it to provide at least as good results as the large lineup.

The relatively small 48-person large lineup was used primarily to make it easier to experimentally check the results of the experiment. It makes little difference to police forces whether they would use a 48-person compared to a 120-person lineup. They have at their disposal thousands of appropriate foils for almost any suspect. On the other hand, researchers do not. The smaller the lineup, the easier it is to acquire the photos.

We must note that Levi's large lineups are not strictly simultaneous ones, since witnesses do not view them all simultaneously. In the 48-person lineup, for example, the photos are divided between four pages of twelve photos each. However, witnesses can move back and forth between the pages. Thus, they can easily compare the photos on each page, and then use their memory to compare the most likely candidate from each page.

Method

Participants: The 82 participants were graduate students and staff of an Israeli University who agreed to participate in a study on memory. About half were male, and ages ranged from twenty to sixty, with a median of thirty.

Design: The design was a between-participant design, in which the between-participant factor was a culprit-present or culprit-absent lineup, with two dependant measures, the verbal response of the witness and the eye tracked response.

Apparatus: The eye tracker used in this experiment was the SMI-RED, a mobile device with a screen and a laptop computer which controlled the experiment, and was considered by the author to be perfectly adequate for the lineup.

Recruitment and Eyewitness Condition

The author visited labs and offices at an Israeli university. The author introduced himself, and asked the occupants whether they would participate in a memory experiment at a later time that would last only about five minutes. If a person agreed, he immediately showed them a video in their office or lab lasting 2 minutes in which the target was seen for 37 seconds, another young-looking male for 22 seconds[1]. He arranged a mutual acceptable time for the experiment, at least an hour later.

----

[1] The video was a natural domestic scene showing a mother diapering her baby in the baby's room, a young-looking male and an older woman sitting in the living room, and the target moving into the living room, sitting down, putting on his shoes, and moving in and out of the room where the mother was diapering the baby.

Procedure

The witnesses were told that they were to view a lineup of four screens of 12 photos to see whether they could identify the target ("the male who moved around"), who may or may not be in the lineup. They could view the lineup as many times as they wanted before making their decision. While viewing the lineup their eyes would be tracked. The eye tracker was then calibrated. When the calibration was satisfactory the lineup was shown.

The lineups

Photos for the lineups were chosen from Levi (2012). All lineup members were young adult males who had dark and short hair, dark eyes, no beard or moustache, and were of medium build. The target also fit this description. The twelve faces of each screen were organized in two lines of six. The four screens were identical for the target-present and target absent lineup, except that the target was placed in the lower left hand corner in the target-present lineup's fourth screen, and replaced with a different photo in the culprit-absent lineup. The order of the four screens was randomly determined for each participant.

Results

Table 1 presents the results for the responses of the witnesses. For target-present lineups there are three possible verbal responses: The target is chosen, some other lineup member (foil) is chosen, or no one is chosen. For these lineups the verbal responses were fairly evenly divided between 18 correct identifications (30.5%), 20 foil identifications (33.3%), and 22 no choices (37.3%). For target-absent lineups, the responses were again fairly evenly divided between 12 incorrect choices (54.5%) and 10 correct rejections (45.5%).

Table 1

Results for verbal and eye response

Target-present	verbal	Eye tracker
Identifications	18 (30.5%)	20 (33.3%)
Foils	20 (33.3%)	40 (66.7%)
No choice	22 (37.3%)	
Target-Absent		
Incorrect choice	12 (54.5%)	
Correct rejection	10 (45.5%)	

The operational definition of an identification with the eye tracker data was as follows. The measure used in this research is "dwell time", the amount of time that the eyes dwelt on any particular photograph. For each of the four screens of photos, the longest dwell time was divided by the next longest one. For target identification, the resulting number had to be the largest for the target than any of the numbers resulting from this calculation for the other three screens. An identification could not of course occur for witnesses whose dwell time for the target was not the largest in the screen where the target occurred. In addition, the ratio between the target and the next largest had to be at least 1.7 (the average ratio for such identifications was 3.6). We found 20 such cases (33.3%), essentially the same as for the verbal responses

On the other hand, we should not, nor need not, be able to distinguish between foil identifications and no choice responses. According to our theory, when witnesses cannot identify the target they compare the lineup members without focusing on any one. Thus, the two categories can be combined. We found a very similar distribution between verbal non-identifications (20, 70.6%) and the combined categories of foil and no choice (40, 66.7%) in the eye tracking data.

According to this same conceptualization, there should be no such thing as either an incorrect identification or a correct rejection in terms of eye movements in target-absent lineups. In all target-absent lineups witnesses should be looking at the lineup members without concentrating on any one in particular.

However, the data do not bear out the predictions of the theory, either in target-present or target absent lineups. Using the same criterion of at least 1.7 to determine an "identification", 33 of the 40 (82.5%) target-present cases that were not target identifications would be termed identifications of some foil. The average ratio was 3.4, very similar to the average ratio of 3.6 for the target identifications. Using the same criterion of 1.7 for the target-absent cases, 19 of the 22 (86.4%) would be termed identifications of some foil. The average ratio was 4.3.

That is, witnesses most often focused on some foil when they did not identify the target in target-present lineups, or could not do so in target absent ones. Adding up these two types of cases, we find that in 52 out of 62 lineups witnesses acted contrary to the relative judgment conceptualization. By the binomial, the probability that so many cases would be contrary to the theory is p< 0.0001 (two-tailed). This experiment did not merely fail to reject the null hypothesis. If found results exactly the opposite of the research hypothesis, very significant statistically.

Another interesting finding is that in 22 cases of the total 82 (26.8%), a foil that fit the criterion of 1.7 was actually verbally chosen by the witness. Finally, in two cases witnesses said that the target was not in the lineup, but by our gaze criterion they had identified him.

Discussion

This experiment had witnesses view a 48-person lineup divided into four screens of twelve photos each. The movements of the witnesses' gaze over each screen were recorded using an eye tracker, and the time the gaze dwelt on each photo was recorded. The hypothesis was that while witnesses identifying the target would spend much of their time gazing at him, witnesses who could not identify him would compare between the lineup members and thus no lineup member would stand out as being viewed substantially longer. As a result, dwell time, the amount of time witnesses dwelt on each photo, might be a superior identification measure to verbal responses.

The hypothesis was strongly disproved by the data. The null hypothesis was rejected in the opposite direction from the theory's prediction. While witnesses did indeed spend much more time gazing at the target when they verbally identified him, they spent as much time gazing at some other lineup member when they could not identify him. Thus, dwell time did not distinguish between correct identifications and incorrect ones.

Therefore, the conceptualization that witnesses compare between lineup members when they cannot identify the culprit (either because their memory is poor when the culprit is present or when he is absent) seems to be incorrect, at least for a 48-person lineup. Rather, the gaze pattern points to witnesses simply focusing on a foil, sometimes mistakenly identifying him. Mansour et al. (2009) concluded the opposite, that their results validate the theory. This is because they found that the vast majority of their witnesses did indeed at least quickly look at all the lineup members. We have posited that a cursory look at lineup members does not validate the relative judgment theory.

Mansour et al. (2009) asked their witnesses whether they had used relative judgment or not on their final trial (each witness viewed a number of lineups). They found that 63% of their witnesses said that they had not used relative judgment, though only 10% had failed to at least glance at each of the lineup members. While verbal responses such as these must always be taken

with a grain of salt, this large discrepancy flies in the face of their conclusion, while dovetails with our interpretation that just glancing at all lineup members does not amount to actually comparing between them.

One attempt to save the relative judgment conceptualization would have the witness comparing each lineup member with a particular one, which then would be gazed at more than any other. (Personal communication Rod Lindsay, June 2014). There are two problems with this explanation. First of all, the pattern of gaze behavior for witnesses who identified the target is virtually identical to those who did not. It seems that we must assume that those who identified him had better memory of him, and therefore were less likely to use relative judgment: they gazed less at other lineup members. Therefore, we should expect that the target would stand out more compared to those with poorer memory who were using relative judgment and comparing between lineup members.

Secondly, 86% of witnesses in target-absent lineups gazed at one member substantially more, yet 46% rejected the lineup. If gazing at one member substantially more is using relative judgment, that would mean that almost half used relative judgment, yet went on to reject that person. That is not what is supposed to happen according to relative judgment. They are supposed to pick him, the person who looks most similar to the target.

One might argue that the relative judgment theory holds true for six-person lineups, but not for 48-person ones. Perhaps. After all, the task of comparing between six people is easier than comparing between 48 people spread out over four screens. On the other hand, the percentage of mistaken choices in culprit-absent lineups has remained at the average of about 50% found in six-person lineups. That high number is explained by the relative judgment conceptualization for the six-person lineup, but not for the 48-person one. An alternate explanation is that as in the 48-person lineup, witnesses with poorer memory in six-person lineups simply mistakenly identify one of the lineup members as the target.

If indeed the 48-person lineup discourages relative judgment, that would be an additional reason for abandoning the sequential lineup, whose main advantage is claimed to do just that (Lindsay & Wells, 1985). In the sequential lineup witnesses view the lineup members one at a time, only once. Thus, the ability to compare between lineup members is severely limited.

It would seem that we might want to look elsewhere for an explanation. Ebbeson and Flowe (2002) posit a criterion shift. Levi (2007b) explains that witnesses, not knowing if the next photo yet to be seen may seem to be the target, wait to see the next photo, and thus fail to identify the one presently seen, who may be the target or the person they would have chosen in a simultaneous lineup. Finally, Levi (under review) has found evidence suggesting that witnesses with partial memory of the target are able to discount at least one of the foils. Then, when they guess among the remaining in the simultaneous lineup, that have a far greater chance of picking the target by chance. These theories also explain the reduction in correct identifications using the sequential lineup

Without being able to count on the gaze pattern of witnesses , we must fall back on the verbal responses. We found that, as often found in simultaneous lineups, about 50% of the witnesses mistakenly chose someone in the target-absent lineups. We have noted that in the common six-person lineup, this would amount to 50/6 = 8.3% mistaken identifications of an innocent suspect. The 84-person lineup saves the day. Only 50/84 = 0.6% of innocent suspects would be identified.

In previous experiments testing large lineups (Levi, 2006b; 2007a; 2012), target identifications never reached the 30.5% found in this study. The most reasonable explanation for this, which occurred in lineups as small as 12, was likely the very difficult lineup event. Levi combined his recruitment of witnesses with his eyewitness event. He visited offices and labs with a student-aged confederate, whose job was to find a mutually acceptable time for the experimental

----

session. In the process the confederate asked for the witness' name and office phone number, but otherwise seemed quite unimportant. Witnesses of course had no idea that they were supposed to remember anything about the interaction, let alone the face of the confederate.

The event of the present study was not easy. The video included four adults and a baby, two of the adults being young males. The target was viewed for only 37 seconds. There were many objects that were videoed, not all in the background. The witnesses knew that they were supposed to remember something in the video, but they were given no clue as to what. In contrast to Levi's studies, however, the learning was not incidental, which explains the higher identification rate. The identification rate is an important issue if we expect police to adopt a large lineup, and a 30% rate may satisfy police departments (see Valentine, Pickering & Darling, 2003 for an extensive summary of real lineup data predating the English video lineup).

The great advantage of large lineups is obvious in reducing the chance of mistaken identifications. In this experiment we have noted that while 54.5% of witnesses chose someone in target-absent lineups the chance of an innocent suspect being chosen is only 54.5/84 = 0.65%. Levi (2012) has tested a 120-person lineup, and found no higher rates of witness choosing in target-absent lineups or lower identification rates in target-present ones. Thus, using a 120-person lineup should result in only 54.5/120 = 0.45% innocent suspects being falsely identified.

Why was the 1.7 ratio chosen as the defining ratio? The ratio of 1.7 was chosen as being the smallest ratio which stood out somewhat. The ratio of 1.6 had three cases in which it occurred twice within the same lineup and four more where the next smallest ratio was 1.5. When 1.7 occurred, on the other hand, the next highest ratio was 1.5. Of course, given the data the exact ratio chosen is of little importance. Any ratio would effect equally target identifications, identifications of foils in target-present lineups, and identification of foils in target absent lineups. Furthermore, 1.7 was the lowest possible ratio defining a focus on the lineup member. We have noted that the average ratios were 3.6, 3.4 and 4.3. That is, the average time that witnesses looked at the person most looked at was more than three times the time the witness looked at the next most looked at person.

As we have noted, constructing 120-person lineups for the police is relatively easy. Police have in their mug shot albums hundreds of thousands of photos to choose from in selecting foils for lineups. Researchers, on the other hand, very rarely have enough photos to construct so large fair lineups. Thus, ironically, police may be unlikely to adopt large lineups because there will be inadequate research on them. Perhaps the solution is to conduct the research with police departments that are willing to contribute the photos, and photograph the target using their standard procedures.

Police departments might also be interested in using eye trackers. We noted that two witnesses announced that the target was not in the lineup though they gazed at him for a long time. Those two witnesses account for 100x2/18 = 11.1% of all the identifications in this experiment. With large lineups, it is quite unlikely that those two cases are a chance finding.

We can understand how this might happen. When witnesses are faced with a decision to either identify someone or declare that the culprit is not in the lineup, they are likely sometimes not to be completely certain that either statement is true. They have to set themselves a criterion (a certain probability) that they can identify someone, below which they do not. Memory can vary from witness to witness, and the degree of certainty can also vary. Witnesses who gazed a long time at the target but failed to identify him probably had a degree of certainty that failed to reach their criterion.

While it is doubtful that courts would put much stock in this gaze behavior, the police in such cases could justifiably devote extra resources to find additional evidence for such a target, who would of course be the suspect.

References

1. *Brace N.* Investigating facial composite construction using eye movements. Proceedings of Sarmac Conference IX, June 27-29, 2011. NYC.

2. *Conners E, Lundregan T, Miller N, McEwen T.* Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial. Washington: U. S. Department of Justice, 1996.

3. *Ebbesen E. B., & Flowe H.* Simultaneous v. sequential lineups: What do we really know? http://www-psy.ucsd.edu/%7eeebbesen/SimSeq.htm

4. *Holmquist K., Nystrom M., Andersson R., Dewhurst R., Jarodzka H., & Van De Weijer J.* Eye Tracking. Oxford: Oxford Press. 2011

5. *Hunter S., & Pike G.* Investigating eye movement for sequential video VIPER line-ups. Proceedings of Sarmac Conference IX, June 27-29, 2011. NYC.

6. *Levi A. M.* Are defendants guilty if they were chosen in a lineup? Law and Human Behavior, 22, 1998. 389-407.

7. *Levi A. M.* An Analysis of Multiple Choices in MSL Lineups, and a Comparison with Simultaneous and Sequential ones. Psychology, Crime, & Law,12, 2006a. 273-285.

8. *Levi A. M.* A Comparison Between Large Simultaneous and MSL Lineups, with Photos Viewed in Sets of Six. In K. Nixon (Ed.) Forensic recall and eyewitness testimony. (pp.91-101) 2006b. London: IA-IP Publishing.

9. *Levi A. M.* Evidence for Moving to an 84-Person Photo Lineup Journal of Experimental Criminology, 3, 2007a. 377-391.

10. *Levi A. M.* Multiple Choices in Large Sequential Lineups with Children and Adults. Cognitive Technology., 12, 2007. 4-10.

11. *Levi A. M.* Much Better than the Sequential lineup: A 120-person lineup. Psychology, Crime & Law, 18, 2012. 631-640.

12. *Levi A. M.* (under review). When the relative judgment theory proved to be false.

13. *Lindsay R. C. L, & Wells G.* Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. Journal of Applied Psychology, 70, 1985. 556-564.

14. *Loftus E. F., Loftus G. R., & Messo J.* Some facts about 'weapon focus'. Law and Human Behavior, 11, 1987. 55-62.

15. *Mansour J. K. , Lindsay R. C. L, Brewer N., & Munhall K. G.* Characterizing visual behavior on a lineup task. Applied Cognitive Psychology, 23, 2009. 1012 - 1026.

16. *Pike G.* Violent, realistic and unexpected staged crimes: Do participant- witnesses behave differently? Proceedings of Sarmac Conference IX, June 27-29, 2011. NYC.

----

17. *Pryke S., Lindsay R. C. L., Dysart J. E., & Dupuis P.* Multiple independent identification decisions: A method of calibrating eyewitness identifications. Journal of Applied Psychology, 89, 2004. 73-84.

18. *Scheck B., Neufeld P., & Dwyer J.* Actual innocence: When justice goes wrong and how to make it right. New York: Signet. 2001.

19. *Sporer S. L.* Eyewitness Identification Accuracy, Confidence, and Decision Times in Simultaneous and Sequential Lineups. Journal of Applied Psychology, 78, 1993. 22-33.

20. *Stebla N., Dysart J., Fulero S., & Lindsay R. C. L.* Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. Law and Human Behavior. 25, 2001. 459-474.

21. *Valentine T., Pickering A., & Darling S.* Characteristics of eyewitness identification that predict the outcome of real lineups. Applied Cognitive Psychology, 17, 2003. 969-993.

22. *Wells G. L.* The psychology of lineup identifications. Journal of Applied Social Psychology, 14, 1984. 89-103

23. *Wells G. L., Small M., Penrod S., Malpass R. S., Fulero S. M., & Brimacombe C.A. E.* Eyewitness identification procedures: Recommendations for lineups and photospreads. Law and Human Behavior, 22, 1998. 603-647.