Eyewitness Testimony

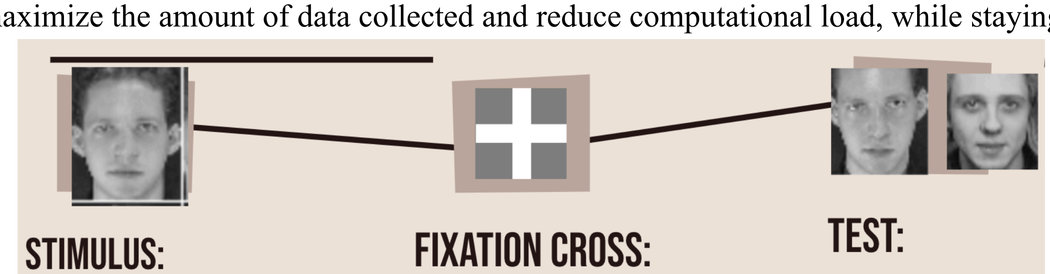
Introduction:

Last year, I watched a criminology documentary depicting the inaccuracies of eyewitness testimony and the large amount of false convictions that result from it. Witnesses of a crime would confidently single out a suspect from a lineup, only to be assertively incorrect. This sparked my curiosity to observe this phenomenon firsthand. I wanted to test my own assumptions about identification accuracy, especially the belief that clear visibility of a suspect's face guarantees correct identification. A point of interest from the documentary was the notion that both unconfident and confident identifications were inaccurate. Thinking in the context of an experiment, I wanted to analyze a primary contributing factor to confidence: the duration of observation. How much would accuracy change if a witness saw the suspect for a short duration in comparison to a longer duration? In a real world context, how much more accurate would a witness be when identifying a suspect in a robbery—where faces are often visible for a longer period of time—versus a hit and run—where observation may be fleeting. Intuitively, I hypothesize that participants would be significantly more accurate in the longer exposure condition than the shorter exposure condition.

Experimental Design:

The experimental design was a within-subjects design, with randomized conditions and 2 Alternative Forced Choice (2AFC) trials. There were 2 conditions: a long exposure condition and a short exposure condition, in which the observer was presented with the suspect for 1.5 seconds and 0.4 seconds, respectively. These durations were chosen through personal trial and error—randomly testing which durations were sufficiently long without being trivial and which durations were sufficiently short without being too difficult. The goal was to simulate real-world observation scenarios where witnesses might have either fleeting or sustained opportunities to view a suspect committing a crime. In total, there were a total of 30 trials: 15 trials under the short exposure condition and 15 trials with the long exposure condition. Over the course of the 30 trials, the order of these conditions were randomized, to mitigate sequence effects. The experiment started with an instruction phase with a sample trial to allow participants to familiarize themselves with the key inputs and structure. Afterward, each trial began with the **stimulus phase**, where the suspect was presented for a randomized duration of either 1.5 seconds or 0.4 seconds. This was followed by a **retention phase** containing a 10 second fixation cross, intending to replicate a period of memory loss/distortion between the time a witness observes a crime and the suspect identification test. The final portion of the trial was a **test phase** where a different picture of the suspect was presented next to an innocent person. In this phase, the position of the suspect is randomized and the participant is asked to identify the suspect position using the [left] or [right] keys. Furthermore, the age, race, gender, expression, lighting, and angle were controlled and forced similarities between the innocent person and suspect for each individual test phase.

This experimental construction poses a possible question: why use a 2AFC design for the trials? Put simply, given the time and resource constraint of this experiment, a 2AFC design would maximize the amount of data collected and reduce computational load, while staying true



to the natural decision making process. In the context of this experiment, the 2AFC design is a viable alternative to a traditional nAFC design with a lineup, due to the decision theory principle of “elimination by aspects”. Proposed by Amos Tyversky, a cognitive and mathematical scientist, this phenomenon postulates that people evaluate options one attribute (or "aspect") at a time, discarding alternatives that do not meet their criteria for that attribute, until they reach two options. For example, a witness would eliminate subjects in a lineup by race, facial hair, gender, and other features until there are two similarly looking subjects left. Comparatively, by presenting only two options, the experiment streamlined this natural decision process, presenting only two similarly looking people that participants can directly compare. Ultimately, this 2AFC design removes the need for exhaustive comparisons across multiple options, allowing for more trials, as participants do not need to comprehensively scan a lineup for every trial. This also reduces participant fatigue and positional bias, leading to more accurate results. While this approach sacrifices some ecological validity, such as the absence of a full lineup, it offers a controlled and efficient means of analyzing the effects of exposure duration on identification accuracy.

Results:

Participants demonstrated a higher overall proportion of correct responses in the long exposure condition compared to the short exposure condition. Specifically, the proportion correct was 72.93% in the long condition and 69.08% in the short condition. Similarly, d-prime (d') values, which reflect sensitivity to distinguishing the suspect from the innocent person, were higher in the long condition (d' = 1.223) than in the

```
Aggregate Condition-Level Stats:

Long Condition:
  Overall Proportion Correct (Hit Rate): 0.7293233082706767
  False Alarm: 0.2706766917293233
  D-prime (FC): 1.222797539339794
  Lambda (FC): -0.045449947737033924
  Log Beta (FC): -0.05557608425596733

Short Condition:
  Overall Proportion Correct (Hit Rate): 0.6907894736842105
  False Alarm: 0.3092105263157895
  D-prime (FC): 1.002484340843997
  Lambda (FC): -0.11236690613903055
```

short condition (d' = 1.002). These results agree with the hypothesis and suggest increased accuracy and higher sensitivity under longer exposure durations.

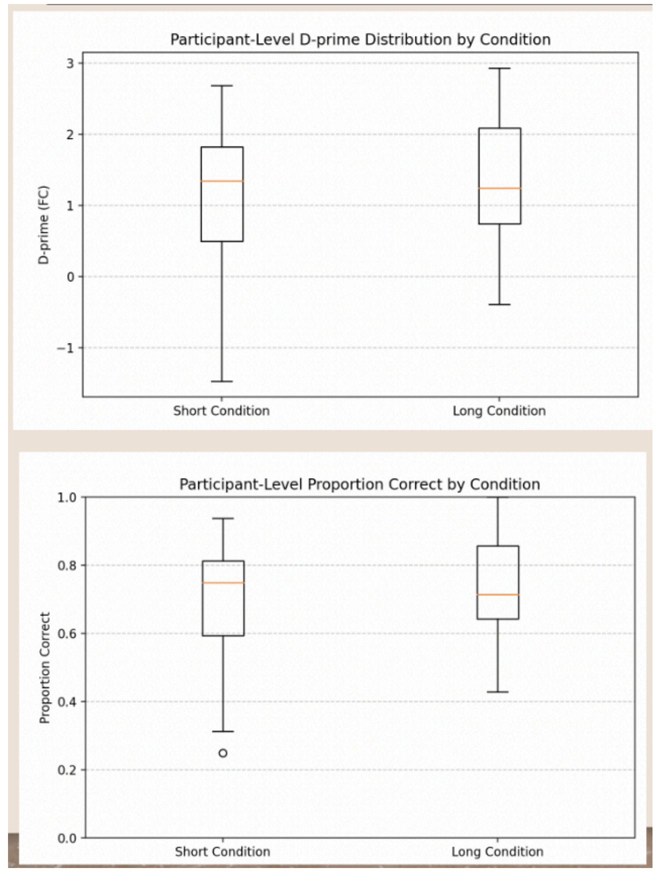
Lambda (λ) and Log Beta were both negative in both conditions, indicating a positional or decision bias favoring the selection of the signal-left (SN) option. This bias was slightly larger in the short condition (λ = -0.113 and Log Beta = -0.113) than in the long condition (λ = -0.045 and Log Beta = -0.056). This could reflect increased uncertainty under shorter exposure durations, leading to strategic adjustments in decision-making, such as defaulting to choosing left.

On an individual performance basis, the data revealed considerable variability in both sensitivity and accuracy across participants and conditions. In the short exposure condition, d' values ranged from -1.469 to 2.684, showcasing a wide disparity in participants' ability to distinguish between the suspect and the innocent individual. For the long exposure condition, d' values spanned from -0.386 to 2.930, suggesting improved but still variable performance with extended observation time. For proportion correct, the short exposure condition showed participant accuracy ranging from 25% to 93.75%, while the long exposure condition ranged from 42.86% to a perfect 100%. These findings suggest that, while longer exposure durations generally improved accuracy, some participants still struggled to perform consistently, even under favorable conditions.

```
Summary of Maximum and Minimum Values:
  Max d' (Short Condition): 2.684469924728554
  Min d' (Short Condition): -1.468988744340383
  Max d' (Long Condition): 2.930467585371046
  Min d' (Long Condition): -0.38593645214015815
  Max Proportion Correct (Short Condition): 0.9375
  Min Proportion Correct (Short Condition): 0.25
  Max Proportion Correct (Long Condition): 1.0
  Min Proportion Correct (Long Condition): 0.42857142857142855
```

To analyze if there was indeed a statistically significant accuracy improvement in the long condition versus the short condition, a paired t-test was conducted to compare d-prime and lambda values between the two conditions. The d-prime paired-test yielded a t-statistic of -0.701

and a p-value of 0.492. For the lambda values, the paired t-test produced a t-statistic of -0.569 and a p-value of 0.576. By the paired t-test, neither of these comparisons were statistically significant, as they exceeded the threshold of 0.05. These results suggest that under this experiment and the conditions tested, exposure duration alone may not have a substantial impact on performance.



From a graphical standpoint, the median d' value and proportion correct were higher in the long exposure condition. This suggests that participants were indeed more accurate when given more time to observe the suspect. In contrast, the short exposure condition demonstrated greater variability in both d' and proportion correct, indicating that participant performance was less consistent under limited viewing durations. This variability may reflect individual differences in memory encoding and decision-making ability under time constraints, as shorter exposure times increase cognitive load and uncertainty. In fact, on an individual basis, theHowever, the plots depict substantial overlap between the two conditions, consistent with the finding of no statistically significant difference.

Overall, the lack of statistically significant results suggests that while exposure duration may play a role in eyewitness accuracy, it is likely one of many interacting factors. Particularly in this experiment, a possible critique could be that the duration of the stimulus or the retention phase delay is not long or short enough to make a big difference. It is probable that longer durations would better depict the impact of encoding time. Furthermore, a possible confounding variable is variation in degrees of similarity between the suspect and innocent person. For any given trial the set of faces could be more similar than other trials, making identification more difficult. Future experiments can take a look at additional variables, such as stress or environmental distractions, to further understand the nuances of eyewitness identification. Regardless of these limitations, this study provides a foundational step in understanding the nuanced relationship between observation duration and decision-making in real-world contexts, such as criminal investigations.

Other literature, such as the study titled “Diagnosing Eyewitness Identifications with Reaction Time-Based Concealed Information Test: The Effect of Observation Time” by Sauerland, Koller, Bastiaens, and Verschuere, highlights the relationship between observation time and the accuracy of eyewitness identifications. This research found that longer exposure times generally improve identification accuracy, suggesting a moderate linear positive correlation between the duration of observation and a witness's ability to correctly identify a suspect. The study looked at both implicit and explicit identification processes in a CIT examination using reaction times and simultaneous photo lineups. They were able to demonstrate that favorable viewing conditions, such as extended observation, significantly enhance face recognition accuracy. However, it also noted that despite the advantages of longer observation times, accuracy remains imperfect, noting the inherent challenges of eyewitness testimony.

Conclusion:

Although eyewitness misidentification makes up 75% of wrongful convictions, it is very often the only lead in criminal investigations. After all, a jury is incredibly convinced when a witness confidently singles out someone they think is the suspect. Therefore, human errors caused by memory distortion raises questions about its admissibility in the judicial system. Understanding these factors is paramount for preventing wrongful convictions. While the findings suggest that longer observation times improve accuracy and sensitivity to some extent (though not statistically significant), they also reveal that accuracy is far from perfect, even under optimal conditions. Indubitably, eyewitness testimony must be scrutinized with caution and further research analyzing further variables such as stress, race, and possession of a weapon must be conducted.