5.17 Social-Cognitive Biases In Simulated Airline Luggage Screening

SOCIAL-COGNITIVE BIASES IN SIMULATED AIRLINE LUGGAGE SCREENING

Jeremy R. Brown & Poornima Madhavan
Old Dominion University
jrbrown@odu.edu pmadhava@odu.edu

Abstract. This study illustrated how social-cognitive biases affect the decision making process of airline luggage screeners. Participants (n = 96) performed a computer simulated task to detect hidden weapons in 200 x-ray images of passenger luggage. Participants saw each image for two (high time pressure) or six seconds (low time pressure). Participants observed pictures of the "passenger" who owns the luggage. The "pre-anchor group" answered questions about the passenger *before* the luggage image appeared, the "post-anchor" group answered questions *after* the luggage appeared, and the "no-anchor group" answered *no* questions. Participants either stopped or did not stop the bag, and rated their confidence in their decision. Participants under high time pressure had lower hit rates and higher false alarms. Significant differences between the pre-, no-, and post-anchor groups were based on the gender and race of the passengers. Participants had higher false alarm rates in response to male than female passengers.

1.0 Visual Search Tasks

The primary goal of visual search tasks is to effectively differentiate critical signal stimuli from irrelevant non-signals (known as *distractors*). There have been various studies looking into different aspects of visual search tasks. Many of the visual search studies focus on visual clutter and its effects on the search task [1, 2, 3,& 4]. Another factor that affects visual search is age [1, 2]. Visual clutter is typically caused by "distractors". Studies by Grahame Laberge, and Scialfa (2004) [1] and McPhee, Scialfa, Dennis, Ho, and Caird (2004) [2] found that as clutter is increased, the time it takes to detect the target also increased. They found that the task increased in perceived difficulty as a consequence of increased clutter. This is because it is harder to recognize an object as the clutter increases [5]. As there are more objects to search through to find a target, the search will take longer and will be less efficient. In some instances, however, detection time can *decrease* with clutter, especially when the clutter causing objects are of a larger size than the target [6]. This is due to attention being drawn to the "empty" space between the clutter causing objects.

In addition to the amount of clutter, search efficiency is affected by what the clutter consists of and its physical similarity or dissimilarity to the target. The more similar the distractor is to the target, in terms of color, brightness, and orientation, the more difficult it is to find the target [3]. Target objects that have multiple colors or textures are harder to detect in a cluttered environment, especially when the clutter is of a similar color or texture to that of the target [7].

The reason visual search tasks are the focus of several researchers is that there are several jobs in the real world that use visual search as the main component of the work such as airport luggage screening. The primary task for airline luggage screening requires the screener to search through an x-ray image and detect a particular dangerous target from the clutter of non-lethal objects. On one level, luggage screening is a simple signal detection task where the screener must differentiate critical signals (or, threat objects) from background noise. However, the detection task is complicated by the fact that on several occasions, the threat object must be detected within an initial glimpse of the x-ray image, spanning just a few seconds.

Airport luggage screening is further complicated by the number and diversity of threat objects that might potentially be embedded in a piece of luggage. All the studies described above have effectively addressed the cognitive aspects of visual search in luggage screening at the level of the individual. However, no study so far has attempted to address extraneous issues (social, cultural, environmental) that might potentially influence screening efficiency over and beyond those that extend beyond simple visual search processes.

The purpose of this study was to examine what effect, if any, variables such as race, age and gender of the passenger have on the screener's decisions to stop the passenger's luggage or not. Computer simulation was used instead of observing actual luggage screeners so that the study could be more controlled than would be possible in the real environment. Simulation also allowed the study to be run using the same luggage images for several student "screeners" allowing comparison between different screeners and luggage images.

1.1 Social –Cognitive Biases

1.1.1 Age

Age bias is a social bias related to a person's age that can have an effect on decision making. Older people often tend to be discriminated against for jobs. Specifically, the belief is that older individuals are not as flexible in their thinking as younger individuals. Therefore a job that requires flexibility would not be a good fit for an older worker, [8] whereas younger people are believed to have more potential for development than the older people [9]. Based on this, younger people may be more likely to be employed as airport luggage screeners, as their thinking must be very flexible to figure out what constitutes a target.

1.1.2 Gender

When one gender is given preferential treatment over the other, it is typically referred to as "gender bias" [10]. Gender bias is pervasive especially in the workplace. When men and women are evaluated for the same type of work male workers are often found to get better rewards for good evaluations compared to female workers; on the flip side, males also receive harsher punishments than females in response to poor evaluations [11]. Research has revealed that performance ratings are more strongly related to promotions for female workers compared to male workers, which suggests that females are held to higher standards than males [12]. For example, in one study wherein men and women were fired from similar jobs, men received more compensation than women [13].

Clearly, gender-related biases play a major role when decisions to hire, promote or fire are made in several job contexts.

1.1.3 Race

Though we would like to think differently, racial bias is still prevalent throughout the world. There have been numerous studies looking at racial bias among police and their decision to shoot or not shoot [14]. In the Correll et al. (2007) [14] study, comparing police to civilians in the same district, civilians were found to be more likely to shoot when shown a minority suspect compared with the police. Both police and civilian participants took longer to react when the White suspect had a gun, and the minority suspect did not have a gun. The researcher concluded that seeing a white person with a gun violated people's expectations leading them to take longer to react; the opposite was true when observing a person of minority race who was perceived as dangerous even without a weapon [14]. The police officers and civilians were White, Black, Native American, and Hispanic so there was a mix of races.

2.0 The Luggage Screening Study

This study was designed to examine whether the social-cognitive biases described above would influence the decision making process in an airport security screening context. What makes this study unique is the focus on social-cognitive biases which differs from existing studies that have focused on either the luggage screening process [15, & 16] or on the decision making made by the luggage screener [17, 18, 19]. This study was designed to examine whether these biases will influence active decision making during the luggage screening process. As mentioned earlier, we implemented a laboratory-based experimental task along with a luggage screening simulation to study this.

3.0 Method

3.1 Participants

Participants were 96 Old Dominion University undergraduates completing the study for class credit. The study took approximately 1 hour to complete, for 1 hour of research credit.

3.2 Materials

Gateway computers were used, which were running Microsoft XP with service pack 2. These computers were used to run a computer simulation of airline luggage screening created by E-prime 2.0.

3.3 Procedure

Participants were randomly assigned to a control group, (n=24), and three experimental groups (n= 72) in a 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, no-anchor, post-anchor) design. Participants filled out an entrance questionnaire prior to running the study. The task was for participants to detect the presence of dangerous objects in x-ray images of passenger luggage. Participants scanned 200 images distributed into two blocks of 100 images each. At the beginning of each block, participants were shown the five targets that they needed to look for in the 100 bags that were to follow. For the experimental groups, the appearance of the luggage image on each trial was preceded by the picture of a random passenger (drawn from a new set of 100, that includes the following races: White, Black, Asian, Middle Eastern, and Hispanic) to whom the bag supposedly "belongs". For each of the experimental groups, half the participants performed the task under high time pressure (2 seconds for each luggage image exposure) and the other half performed under low time pressure (6 seconds for each luggage image exposure). After deciding whether to pass the bag or not, participants rated their confidence in their decision on a five point scale.

Participants in the *pre-anchor group* (n = 24) were first required to answer two statements about the passenger *before* the x-ray image appears. After answering the statements, they clicked "next" and the x-ray image was brought up onto the screen, after which, they rated their confidence on their decision of whether or not to pass the bag. The two statements that were used were statement #1:*"I think this person is attractive"* and statement #2 *"I will most likely stop this person's luggage."* These two statements appear to be the most powerful indicators of the existence of such cognitive biases.

For participants in the *no-anchor group* (n = 24), after 4 seconds of the passenger appearing the x-ray image of a bag appeared beside the passenger. These participants were not required to answer any questions about the passengers, but they still rated their confidence on their decision to pass or not pass the luggage.

For the *post-anchor group* (n = 24), the program ran the same experimental procedure as for the no-anchor group. However, participants were required to answer the two statements answered by the pre-anchor group about each passenger *after* the participant has chosen whether or not to pass the bag and rated their confidence in that decision. Once they have answered the questions and clicked "next", the next picture of a passenger appeared. This procedure continued until the end of the trial block.

A control group (n = 24) performed the screening task alone without observing the pictures of passengers. Of these 24 participants, 12 participants performed under high time pressure and the other 12 performed under low time pressure. This group served as a baseline for performance under each level of time pressure without the additional anchoring information provided by the passengers' pictures.

The base rate for the targets was 50% for all groups. Participants were not informed about the base rate. At the end of each trial, participants received feedback in the form of a text message, telling them whether they made a correct decision or not. Also they received a cumulative percent correct score shown after each decision to pass or not pass the bag. At the end of the experiment, participants filled out a final "task knowledge" questionnaire. The participant with the highest score for their experiment session received a piece of candy as a prize.

4.0 Data Analysis

The data was analyzed for normality. If normality is violated, box plots were used to examine which sections of the data were outliers, and the outliers were brought to 2 standard deviations away from the mean. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was ran for each dependent variable. For the interactions that were significant, a mixed measures ANOVA was run, followed by paired t-tests with a bonferroni correction.

The dependent variables of interest were:

- Hit rate - the probability of correctly detecting a target.
- False alarm rate - probability of an incorrect detection when there was no target
- Sensitivity (d') – the perceptual ability to differentiate between a target and non-target.
- Response criterion setting (c) – the propensity to generate "yes" or "no" responses.

The data analytic strategy was based on a two-pronged approach. We used hit rate and false alarm rate as pure performance measures which directly measure a participant's performance on the task. In addition, we used the signal detection variables of sensitivity and response criterion setting to understand the decision making processes that drive performance (resulting in hit and false alarms).

5.0 Results

Due to the complexity of the experimental design, the study was broken up into two different sets of variables. Hit rate and false alarm rate are grouped under "performance analysis", and sensitivity and response criterion setting are grouped under "signal detection analysis".

5.1 Performance Analysis

5.1.1 Hit Rate

All "p" values below .05 are statistically significant. The hit rate data was normally distributed with no outliers, therefore no data cleaning was necessary.

A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was used to analyze the hit-rate data. The mixed measures ANOVA revealed that there was a significant main effect for time pressure ($F(1,66) = 56.18$, $p \leq .001$, $\eta^2 = .46$). Participants under low time pressure had higher hit rates ($M = .82$, $SE = .01$) than the participants under high time pressure ($M = .69$, $SE = .01$). All other main effects and interactions were statistically non-significant.

5.1.2 False Alarm Rate

The data set was not normally distributed, and the box plots revealed 12 outliers, which were brought in to within 2 standard deviations of the mean. This made the data set normally distributed. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA, similar to that used for the Hit Rate analysis, was used to analyze the False Alarm Rate data. The results of the ANOVA revealed that there was a significant main effect for passenger gender ($F(1, 66) = 7.81$, $p = .007$, $\eta^2 = .11$), and time pressure ($F(1, 66) = 10.80$, $p = .002$, $\eta^2 = .14$). Participants had a significantly higher false alarm rate for male passengers ($M = .16$, $SE = .01$) than they did for the female passengers ($M = .13$, $SE = .01$). Participants under high time pressure ($M = .19$, $SE = .02$) had significantly more false alarms than did the participants under low time pressure ($M = .11$, $SE = .02$). All other main effects and interactions were statistically non-significant.

5.2 Signal Detection Analysis

5.2.1 Sensitivity: d'

Sensitivity, also known as discriminability index, is a measure of how far apart the signal and noise curves are for an individual (Heeger, D., 1997). In other words, this implies that the more the signal (or, target) stands out from back ground clutter, the easier it will be for the human to locate the target. So, in this experiment, higher sensitivity implies that it was easier for the participant to distinguish the target from non-targets. Specifically, the higher the sensitivity, the better was the detection performance.

The sensitivity data was normally distributed with no outliers, therefore no data cleaning was necessary. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was used to examine the data obtained for sensitivity. The main effect of time pressure ($F(1, 66) = 47.34$, $p \leq .001$, $\eta^2 = .418$) and the interaction between passenger gender, passenger race, and anchor ($F(8, 264) = 3.34$, $p = .001$, $\eta^2 = .092$) were both significant. Under low time pressure ($M = 2.23$, $SE = .07$) participants had higher sensitivity than did participants under the high time pressure ($M = 1.54$, $SE = .07$).

To further analyze the relationship between passenger gender and passenger race within each anchor group, a 2 (gender) X 5 (race) mixed measures ANOVA was run within each of the anchor groups and is described in the following sections.

5.2.1.1 Pre-anchor

All of the main effects were non-significant which include the following: passenger gender and passenger race. The interaction between passenger gender and passenger race was significant ($F(4, 92) = 2.863$, $p = .028$, $\eta^2 = .102$). Sphericity was violated, and by using the Greenhouse-Giesser (p=.063), Huynh-Feldt (p=.057), and the Lower Bound (p=.104) correction the interaction became statistically non-significant. All of the other interactions were found to be non-significant which include the following: passenger gender by time pressure, passenger race by time pressure, and passenger gender by passenger race by time pressure.

5.2.1.2 No-anchor

All of the main effects were non-significant which include the following: passenger gender and passenger race. The only interaction that was found to be significant was the interaction between passenger gender and passenger race ($F(4, 92) = 2.621$, $p = .04$, $\eta^2 = .102$). Sphericity was violated and by using the Greenhouse-Giesser (p=.048), and Huynh-Feldt (p=.04) correction the interaction was still statistically significant. However, using the Lower bound (p=.119) correction rendered the interaction statistically non-significant. All of the other interactions were found to be non-significant including the following: passenger gender by time pressure, passenger race by time pressure, and passenger gender by passenger race by time pressure.

The only statistically significant difference between male and female passengers was between the White passengers; participants had higher sensitivity for detecting targets when the passengers were male compared to female (male: $M = 1.87$, $SE = .17$; female: $M = 1.45$, $SE = .16$; $t = 2.786$, $p = .011$).

5.2.1.3 Post-anchor

All of the main effects and interactions were non-significant.

5.2.2 Response Criterion Setting: c

Response Criterion Setting is the propensity to generate "yes" or "no" responses. This means that the human sets an arbitrary threshold or "cutoff point" for responding; when the signal to noise ratio is perceived as being above this level, the participant will indicate a target is present. Likewise, if the ratio is perceived as being below this threshold, they will indicate that a target is not present (Heeger, D., 1997). Typically, if the participant sets his/her response criterion high such that the criterion setting is high or positive, responding is said to be conservative. This means than the participant has a propensity to say "no" more often than "yes". The opposite occurs when a participant sets his/her response criterion low. In such cases, responding is said to be more liberal; this will result in low or negative criterion settings and a general tendency to say "yes" more frequently than "no".

The data set was not normally distributed, and the box plots revealed 12 outliers, which were brought in to within 2 standard deviations of the mean. This made the data set normally distributed. A 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA was used to analyze the response criterion setting data. The ANOVA indicated a significant main effect of passenger race ($F(4, 264) = 8.48$, $p \leq .001$, $\eta^2 = .114$) and an interaction between passenger gender, time pressure, and anchor ($F(2, 66) = 3.50$, $p = .036$, $\eta^2 = .096$). Participants had significantly more conservative response criteria for passengers of Hispanic race ($M = 1.19$, $SE = .09$) compared to all the other races (White $M = .85$, $SE = .05$, $t = 3.97$, $p \leq .001$; Black $M = .83$, $SE = .05$, $t = 4.33$, $p \leq .001$; Asian $M = .82$, $SE = .05$, $t = 4.35$, $p \leq .001$; Middle Eastern $M = .92$, $SE = .06$, $t = 3.14$, $p = .002$). This means that participants were less likely to say there was a target present when confronted with a Hispanic passenger relative to passengers of other races.

To further examine criterion settings within anchor groups, a 2 (gender) X 2 (time pressure) mixed measures ANOVA was run within each of the anchor groups described below.

5.2.2.1 <u>Pre-anchor and Post-anchor</u>

All of the following main effects were non-significant: passenger gender and time pressure. The interaction between passenger gender and time pressure was found to be non-significant as well.

5.2.2.2 <u>No-anchor</u>

For this group all the main effects were non-significant which include passenger gender and time pressure. The interaction between passenger gender and time pressure ($F(1, 22) = 8.391$, $p = .008$, $\eta^2 = .276$) was found to be statistically significant. One tailed *t* tests were used for post hoc analysis of the interaction. The t-tests revealed that there was a non-significant difference for participants' response criterion setting for the male versus female passengers under low time pressure. However, under high time pressure criterion setting for male passengers ($M = 1.14$, $SE = .71$) was significantly higher than for female passengers ($M = .84$, $SE = .70$, $t = 2.18$, $p = .036$).

6.0 Discussion

Most luggage screening studies to date have focused on either mechanics of the luggage screening process [15, 16] or on the decision making of luggage screeners [17, 18, & 19]. What has seldom been addressed in these studies, in particular the decision making studies, is a consideration of extraneous factors, namely social-cognitive variables, that can affect the decision making process. One of these factors is the passenger himself/herself, and any biases the screener may have towards the passengers. The purpose of this study was to examine whether such social-cognitive biases as gender bias, and racial bias would influence decision making during the luggage screening process. We were also interested in examining the role of time pressure, and if the screening process would be affected by decision heuristics such as anchoring.

6.1 Role of anchoring

While time pressure played a significant role in the results, we found that anchoring also played a significant role in impacting decision making. Anchoring is the tendency for decision makers to focus on one particular piece of information and use that to base subsequent decisions [20]. The anchoring heuristic works by giving people a reference point to help them make a decision. For example, in an early experiment on anchoring, when asked a question, "is the percentage of African countries in the United Nations greater than or less than a 25 percent?" [20] Participants generally used the "25 percent" to base their judgment of exactly what percentage of African countries is in the United Nations. This worked even when the percentage was randomly selected in front of the participant. In general, if an anchor is present, the anchor can influence the decision making process of a participant, and therefore influence overall performance.

In this study, the "anchor" was a series of questions drawing attention to the passenger to whom the luggage belonged. Results revealed that when participants had the anchor, either before (pre-anchor) or after (post-anchor) they saw the luggage image, it appeared to suppress rather than enhance the social-cognitive biases, relative to the participants in the no-anchor group. The results also revealed significant interactions between cognitive anchoring and race and gender of passengers on performance. Contrary to our initial expectations and hypothesis, this anchoring effect was particularly salient when time pressure was low and participants had more time to 'attend to' the passengers.

The results suggest that participants used their personal biases as 'anchors' to help in the decision making process, particularly when they had time to pay more attention to passengers. Research has revealed that minority races, such as Hispanics, have been associated with negative behavioral connotations. For instance studies of police officers and their decisions to shoot or not shoot [14], have demonstrated that police were more likely to shoot suspects of minority races even when they did not have a gun. The higher hit rate associated with the Hispanic male passengers in our study could possibly be due to the interaction of these social-cognitive biases. Based on the surmise that the participant already had a negative association with male members of minority races, it is possible that they were more suspicious of the two passenger categories (men and minority races) during the luggage screening process. Therefore, when searching through the x-ray image, they perhaps used gender and race as decision heuristics, paid more attention to the items in bags that were accompanied by male passengers of Hispanic race, and detected the targets more accurately when they were indeed present. This actually suggests a potential benefit of social-cognitive biases in this instance! However, it must be noted that this effect was only observed under conditions of low time pressure when there was ample time to attend to the bags.

The existence of social-cognitive biases in detection behavior is supported, albeit in a slightly different manner, by the false alarm analysis as well. Similar to the effects found in the hit rate data, male Hispanic passengers had a higher false alarm rate associated with them than female Hispanic passengers. Interestingly, the false alarm effect was found under conditions of high time pressure rather than low time pressure. This indicates the negative effects of social-cognitive biases. Although target detection was benefited to an extent due to anchoring under low time pressure, high time pressure led to negative effects in the form of higher false alarms.

Similar effects for racial bias were found in participants' criterion settings wherein participants had a more conservative response criterion setting for certain passenger races. This means that participants were more conservative and had to have a higher subjective evidence of a target being present before they would indicate that one was present. This is very interesting since we have already seen in the false alarm rate data that participants also had a higher false alarm rate for the male Hispanic passengers compared to the other races of passengers. At first glance the conservative criterion setting for Hispanic passengers appears to contradict the finding that participants stopped luggage more (i.e., said "target present" more) in response to these passengers. Is it possible that participants' lower response criterion for the female Hispanic passengers relative to male Hispanic passengers has raised the criterion setting for the Hispanic passengers overall, although this is not evident in a statistically significant difference between the male and female Hispanic passengers per se. As hypothesized, the participants had higher false alarm rates for minority passengers than they did for the White passengers.

As hypothesized, participants had a higher false alarm rate when the passenger was male which would lead them to being stopped more. Also the interaction between passenger gender and time pressure for the no-anchor group was an interesting indication of how not providing an anchor significantly impacted performance more than providing anchors in this study. When time pressure was low, participants had a more liberal response to the male passengers thereby stopping the luggage belonging to male passenger more often. Conversely, participants had a more conservative response towards the female passengers, thereby stopping their luggage with lower frequency than for male passengers. Surprisingly, the opposite became true under high time pressure; participants had a higher, more conservative response to the male passengers, while they had a more liberal response to the female passengers. It is possible that when participants had time to think about the passenger and the luggage, as in the case of the low time pressure group, their biases against male passengers were mitigated to an extent leading them to become more conservative. The opposite might be true for female passengers wherein the index of suspicion possibly increased with the availability of more time to scan the image.

7.0 Conclusions

The results of this research have demonstrated how social-cognitive biases affect people in the real world and how they can subsequently impact the luggage screening process and eventually national security. Through the use of computer simulation we have shown that social-cognitive biases actually do have an effect on the detection of anomalies during luggage screening wherein decision makers use these inherent biases as decision heuristics, particularly under conditions of time pressure. Clearly, such biases would be difficult to detect through mere observation of screening processes at airports. Hence, the use of behavioral experimental and computer simulation is invaluable in such sensitive contexts.

Most importantly, our results revealed a clear relationship between decision making process and performance. Through the use of both signal detection variables and performance variables in our analyses, we are able to draw conclusions not just about the impact of social-cognitive variables on performance, but also the processes that led to the observed behaviors. This is especially important in the current security conscious world we live in and for training of personnel for optimal decision making that is free of biases and prejudices. An associated goal of this research is to the design community for improving the design of imaging equipment and luggage screening stations.

8.0 References

[1] Grahame, M., Laberge, J., & Scialfa, C. T. (2004). Age differences in search of web pages: The effects of link size, link number, and clutter. *Human Factors, 46,* 385-398.

[2] McPhee, L. C., Scialfa, C. T., Dennis, W. M., Ho, G., & Caird, J. K. (2004). Age differences

in visual search for traffic signals during a simulated conversation. *Human Factors, 46,* 674-685 [10] Baker, K., Craddock, A., &Orwig, A. (2002) *Gender Bias*. An educator's guide to access issues. Retrieved July 19, 2008. http://www.ed.uiuc.edu/wp/access-2002/gender_bias.htm.

[3] Verghese, P., & McKee, S. P. (2004). Visual search in clutter. *Vision Research, 44,* 1217-1225.

[4] Nagy, A. L., Neriani, K. E., & Young, T. L. (2005). Effects of target and distractor heterogeneity on search for a color target. *Vision Research, 45*, 1885-1899.

[5] Bravo, M. J.,& Farid, H. (2006). Object recognition in dense clutter. *Perception & Psychophysics, 68*, 911-918.

[6] Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception, 33*, 643-652.

[7] Neider, M. B. (2007). Set-size effects during visual search in realistic scenes. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 68*, 655. Abstract retrieved June 15, 2008 from PsychINFO Database.

[8] Craik, F. I. M. (2002). Human memory and aging. In Psychology at the turn of the millennium, Vol. 1: Cognitive, biological, and health perspectives (pp. 261-280). Bäckman, Lars; von Hofsten, Claes; Hove, England: Psychology Press/Taylor & Francis.

[9] Diekman, A. B., & Hirnisey, L. (2007). The effect of context on the silver ceiling: A role congruity perspective on prejudiced responses. *Personality and Social Psychology Bulletin, 33,* 1353-1366.

[10] Baker, K., Craddock, A., &Orwig, A. (2002) *Gender Bias*. An educator's guide to access issues. Retrieved July 19, 2008. http://www.ed.uiuc.edu/wp/access-2002/gender_bias.htm.

[11] McKay, S., & Tate, U. (2001). Student attitudes regarding gender bias in performance evaluations of salespeople. *Journal of Business and Psychology, 16,* 249-258.

[12] Lyness, K. S., & Heilmann, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology, 4,* 777-785.

[13] Rollings-Magnusson, S. (2004). Gender implications of wrongful dismissal judgments in Canada. *Canadian Review of Sociology and Anthropology, 41,* 27-45.

[14] Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92,* 1006-1-23.

[15] Hilscher, M. B. (2005). Performance implications of alternative color-codes in airport X-ray baggage screening. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*, 3449. Abstract retrieved June 20, 2008 from PsychINFO Database.

[16] McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. C., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science, 15,* 302-306.

[17] Brown, J. R., & Madhavan, P. (2008). Effects of target base rates on visual search performance: A comparison of two scenarios. *Proceedings of the Human Factors and Ergonomics Society 52nd annual meeting*, 333-337.

[18] Madhavan, P., Gonzalez, C., & Lacson F. C. (2007). Differential base rate training influences detection of novel targets in a complex visual inspection task. *Proceedings of the Human Factors and Ergonomics Society 51st annual meeting*, 392-396.

[19] Parasuraman, R., Warm, J. S., & Dember, W. N. (1987). Vigilance: Taxonomy and utility. In L.S., Mark, J.S., Warm, & R. L. Huston (Eds.), *Ergonomics and human factors* (pp. 11-32).

[20] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124-1131.

SOCIAL-COGNITIVE BIASES IN SIMULATED AIRLINE LUGGAGE SCREENING

Jeremy Brown
Advisor:
Poornima Madhavan

Introduction

- Previous studies
  - X-ray color[1]
  - Search Patterns[2]
  - Textual/Spatial Cuing[3]
- Influence of Social-Cognitive Biases
  - Implicit Attitudes Test[4].

1. Ghylin, Schwaninger, Drury, Redford, Lin, and Batta, 2008
2. McCarley, Kramer, Wickens, Vidoni, & Boot, 2004
3. Wiegmann, McCarley, Kramer, and Wickens, 2006
4. Greenwald, McGhee, and Schwartz, 1998

Visual Search Tasks

- goal of visual search tasks
- Effects of Visual Clutter
  - Detection time and perceived difficulty[5]
  - Object recognition[6]

5. McPhee, Scialfa, Dennis, Ho, & Caird, 2004; Grahame, Laberge, & Scialfa, 2004
6. Bravo & Farid, 2006

Time Pressure

- Change in Response Pattern
  - Chess[7]
  - Word Recognition[8]
  - Simulation: Trucking game[9]

7. Harreveld, Wagenmakers, & van der Maas, 2007
8. Light, Chung, Pendergrass, & Van Ocker, 2006
9. Betsch, Fielder, and Brinkman, 1998

Social-Cognitive Biases

- Age
  - Flexibility[10]
  - Potential[11]
  - Facial Recognition[12]
- Gender
  - Rewards and punishments[13]
  - Performance rating[14]

10. Craik, 2002
11. Diekman, & Hirnsey, 2007
12. McKay, & Tate, 2001
13. Anastasi, & Rhodes, 2006
14. Lyness, & Heilmann, 2006; Rollings-Magnusson, 2004

Social-Cognitive Biases

- Race
  - Shoot/Don't shoot scenarios[15]
- Impact of Social-Cognitive Biases on Luggage Screening
  - Young Screeners vs. Old Screeners
  - Males vs. Females
  - Screeners race vs. Passengers race

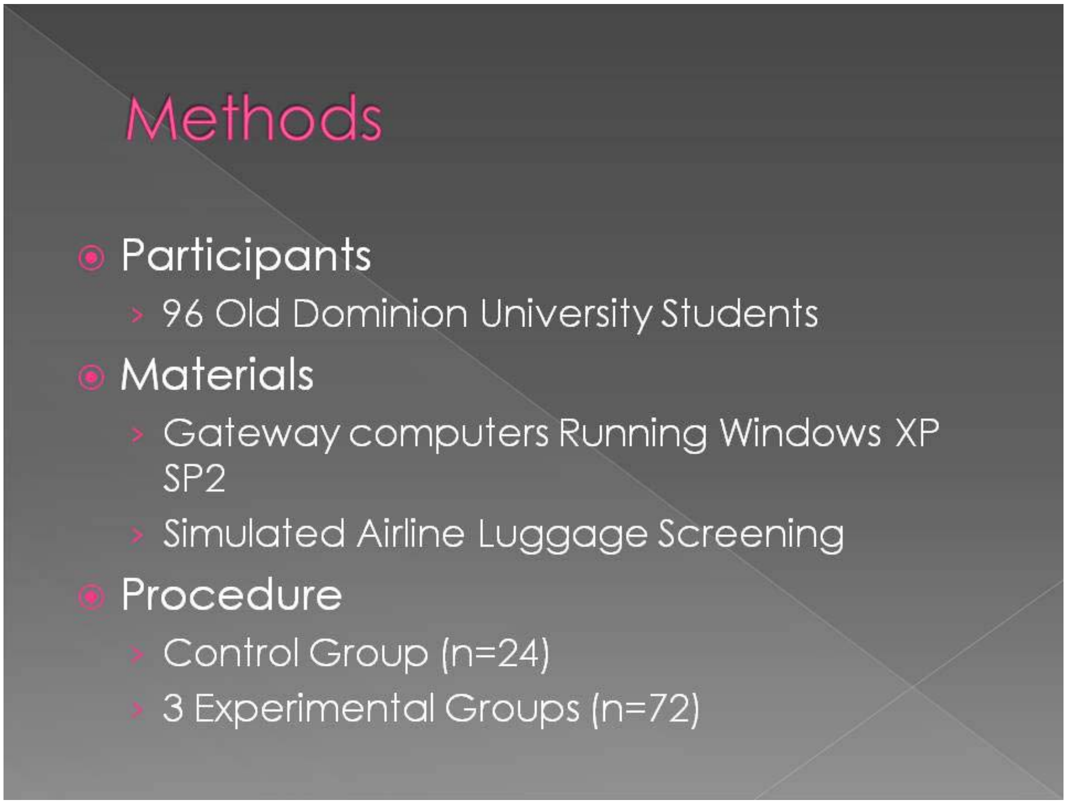
15. Correll, Park, Judd, Wittenbrink, Sadler, & Keesee, 2007; Plant, & Peruche, 2005

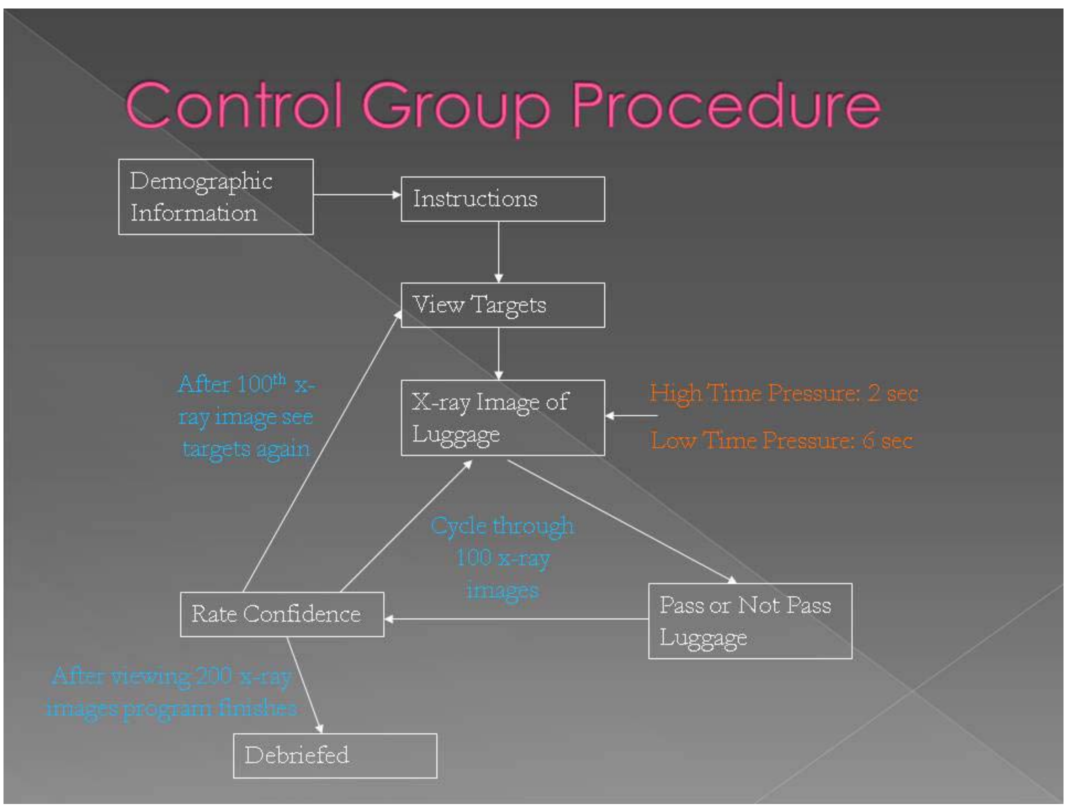
The Luggage Screening Study

- Purpose
  - To examine if Social-Cognitive Biases would influence active decision making

Hypothesis

- Time Pressure and Anchor
  - High time pressure vs. low
  - Pre-, No-, and Post-anchor
- Passenger Gender and Race
  - Male passengers will be stopped more
  - Minority races stopped more


Methods
Participants
96 Old Dominion University Students
Materials
Gateway computers Running Windows XP SP2
Simulated Airline Luggage Screening
Procedure
Control Group (n=24)
3 Experimental Groups (n=72)

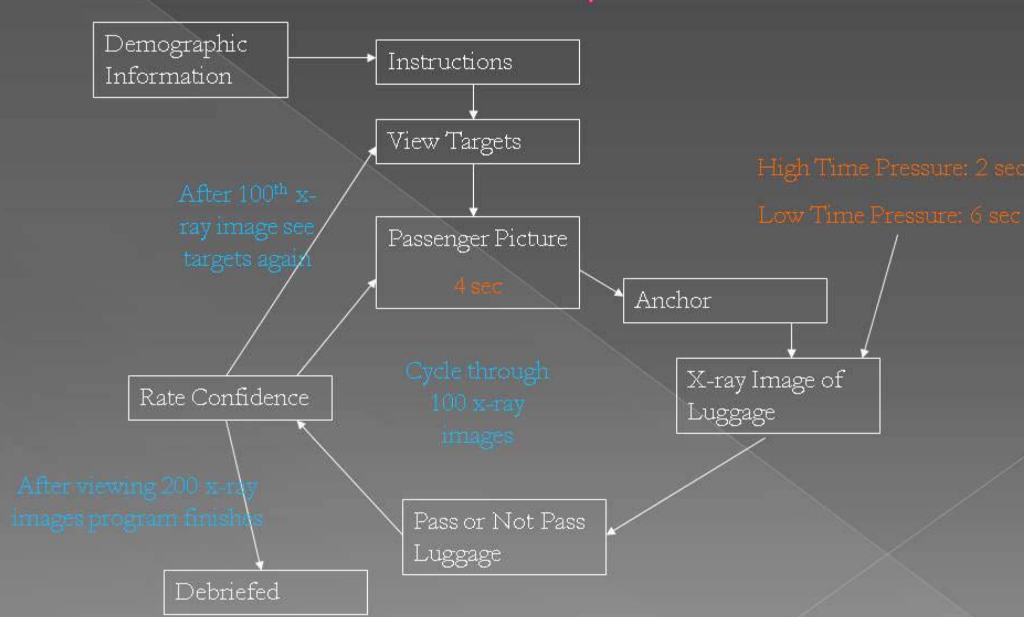

Control Group Procedure
Demographic Information
Instructions
View Targets
X-ray Image of Luggage
High Time Pressure: 2 sec
Low Time Pressure: 6 sec
After 100th x-ray image see targets again
Cycle through 100 x-ray images
Pass or Not Pass Luggage
Rate Confidence
After viewing 200 x-ray images program finishes
Debriefed

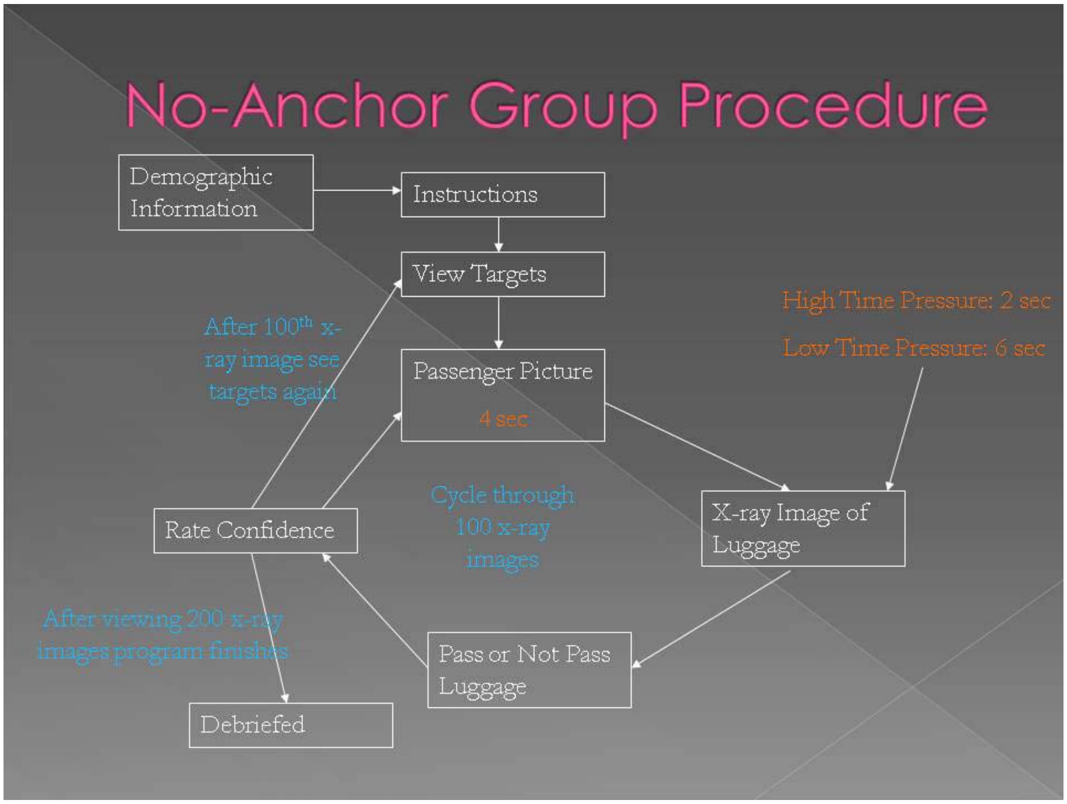
Experimental Groups

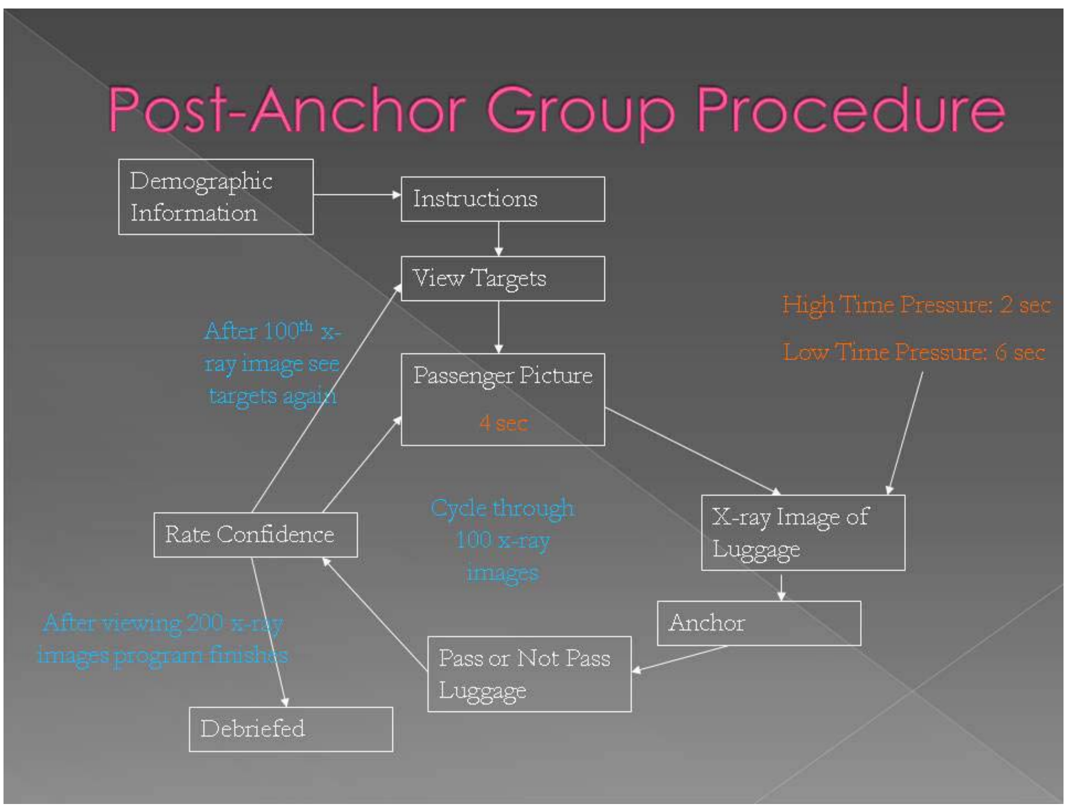
- Anchor
  - *"I think this person is attractive."*
  - *"I will most likely stop this person's luggage."*

  - Pre-anchor (n=24)
  - No-Anchor (n=24)
  - Post-Anchor (n=24)

Pre-Anchor Group Procedure



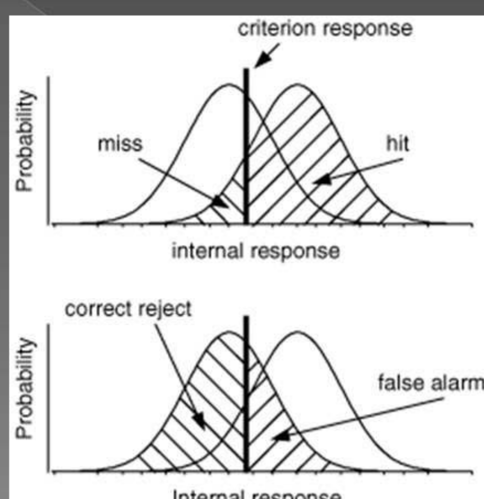

No-Anchor Group Procedure
Demographic Information
Instructions
View Targets
Passenger Picture
4 sec
High Time Pressure: 2 sec
Low Time Pressure: 6 sec
After 100th x-ray image see targets again
Cycle through 100 x-ray images
X-ray Image of Luggage
Rate Confidence
After viewing 200 x-ray images program finishes
Pass or Not Pass Luggage
Debriefed


Post-Anchor Group Procedure
Demographic Information
Instructions
View Targets
Passenger Picture
4 sec
High Time Pressure: 2 sec
Low Time Pressure: 6 sec
After 100th x-ray image see targets again
Cycle through 100 x-ray images
X-ray Image of Luggage
Rate Confidence
Anchor
After viewing 200 x-ray images program finishes
Pass or Not Pass Luggage
Debriefed

Variables of Interest

- Hit rate - the probability of correctly detecting a target.
- False alarm rate - probability of incorrectly saying there is a target present, when there is no target
- Sensitivity (d') – the perceptual ability to differentiate between a target and non-target.
- Response criterion setting (c) – the propensity to generate "yes" or "no" responses.

Sensitivity (d')
Response Criterion Setting



Results

- Performance Analysis
  - Hit Rate
  - False Alarm Rate
- Signal Detection Analysis
  - Sensitivity (d')
  - Response Criterion Setting (c)

Hit Rate

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA
- Time Pressure: $F(1,66) = 56.18$, $p \leq .001$, $\eta 2 = .46$

Hit Rate

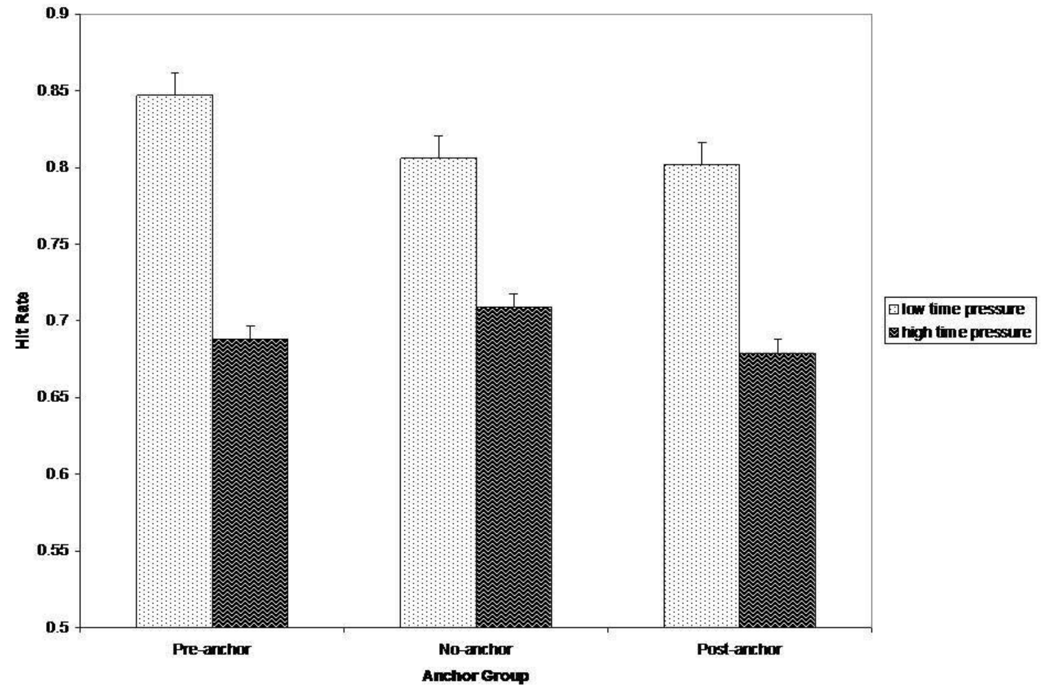
- Non-significant main effects and interactions

Source				
Main Effects	**df**	***F***	**Sig**	**η^2**
Passenger Gender	1, 66	0.66	0.42	0.01
Passenger Race	4, 264	0.713	0.58	0.011
Interactions	**df**	***F***	**Sig**	**η^2**
Passenger Gender X Time pressure	1, 66	0.17	0.68	0.003
Passenger Gender X Anchor	2, 66	0.16	0.86	0.005
Passenger Gender X Time Pressure X Anchor	2, 66	1.01	0.37	0.03
Passenger Race X Time Pressure	4, 264	0.12	0.98	0.002
Passenger Race X Anchor	8, 264	0.96	0.47	0.03
Passenger Race X Time Pressure X Anchor	4, 264	0.71	0.68	0.02
Passenger Gender X Passenger Race	4, 264	1	0.41	0.015
Passenger Gender X Passenger Race X Time Pressure	4, 264	0.67	0.62	0.054
Passenger Gender X Passenger Race X Anchor	8, 264	1.89	0.062	0.054
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	1.71	0.095	0.05

Hit Rate: Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
  - High Time pressure Analysis
    - All main effects and interactions ns
  - Low Time pressure
    - Statistically Significant interaction between passenger gender, passenger race, and anchor
    - $F(8, 132) = 2.071$, $p = .043$, $\eta 2 = .112$

Hit Rate: Anchor Groups



False Alarm Rate

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA
- passenger gender
  - $F(1, 66) = 7.81$, $p = .007$, $\eta 2 = .11$
- time pressure
  - $F(1, 66) = 10.80$, $p = .002$, $\eta 2 = .14$

False Alarm Rate

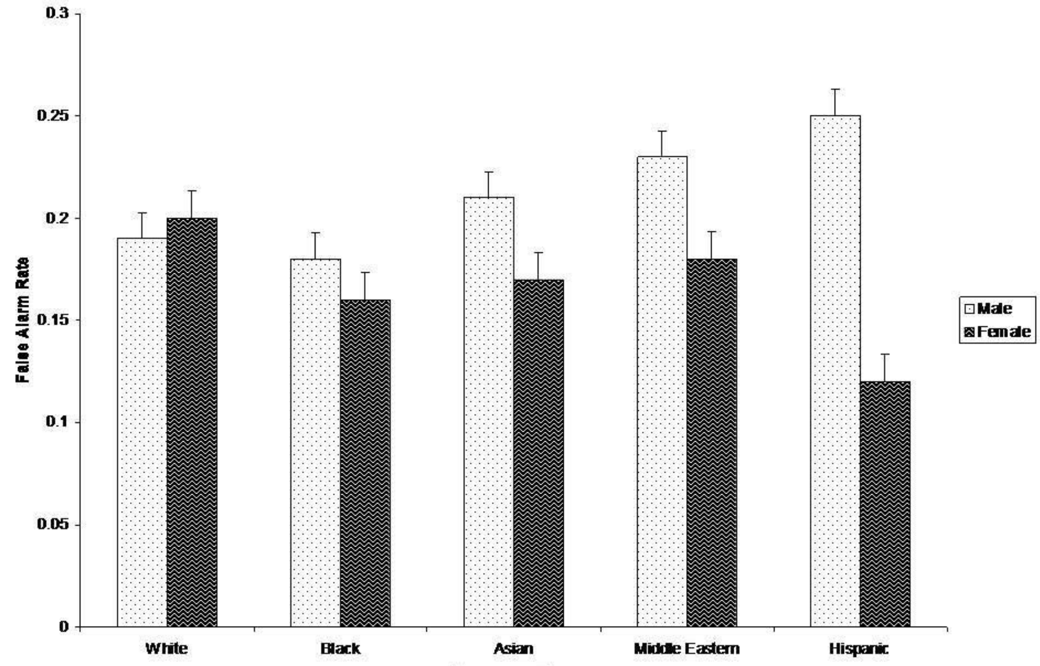
- ns main effects and interactions

Source				
Main Effects	**df**	***F***	**Sig**	**η^2**
Passenger Race	4, 264	0.273	0.9	0.004
Anchor	2, 66	2.01	0.14	0.057
Interactions	**df**	***F***	**Sig**	**η^2**
Passenger Gender X Time pressure	1, 66	1.93	0.17	0.03
Passenger Gender X Anchor	2, 66	0.133	0.88	0.004
Passenger Gender X Time Pressure X Anchor	2, 66	0.171	0.84	0.005
Passenger Race X Time Pressure	4, 264	0.454	0.77	0.007
Passenger Race X Anchor	8, 264	0.201	0.99	0.006
Passenger Race X Time Pressure X Anchor	4, 264	0.636	0.75	0.019
Passenger Gender X Passenger Race	4, 264	1.64	0.16	0.024
Passenger Gender X Passenger Race X Time Pressure	4, 264	1.71	0.15	0.025
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	0.306	0.96	0.009

False Alarm Rate: Time Pressure Groups

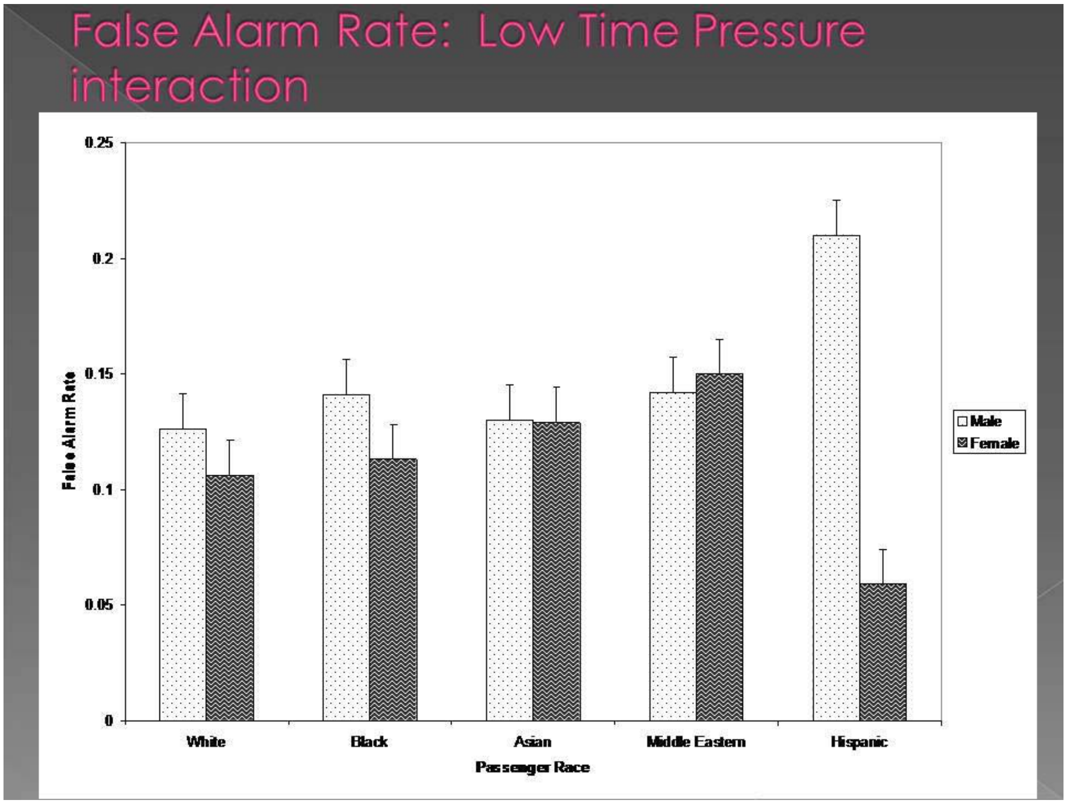
- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
- High Time Pressure
  - gender
    - $F(1, 33) = 8.395$, $p = .007$, $\eta 2 = .20$
  - gender by race interaction
    - $F(4, 132) = 2.430$, $p = .051$, $\eta 2 = .07$
- Low Time Pressure
  - passenger gender, passenger race, and anchor
    - $F(8, 132) = 2.03$, $p = .05$, $\eta 2 = .11$

False Alarm Rate: High Time Pressure



False Alarm Rate: Low Time Pressure interaction

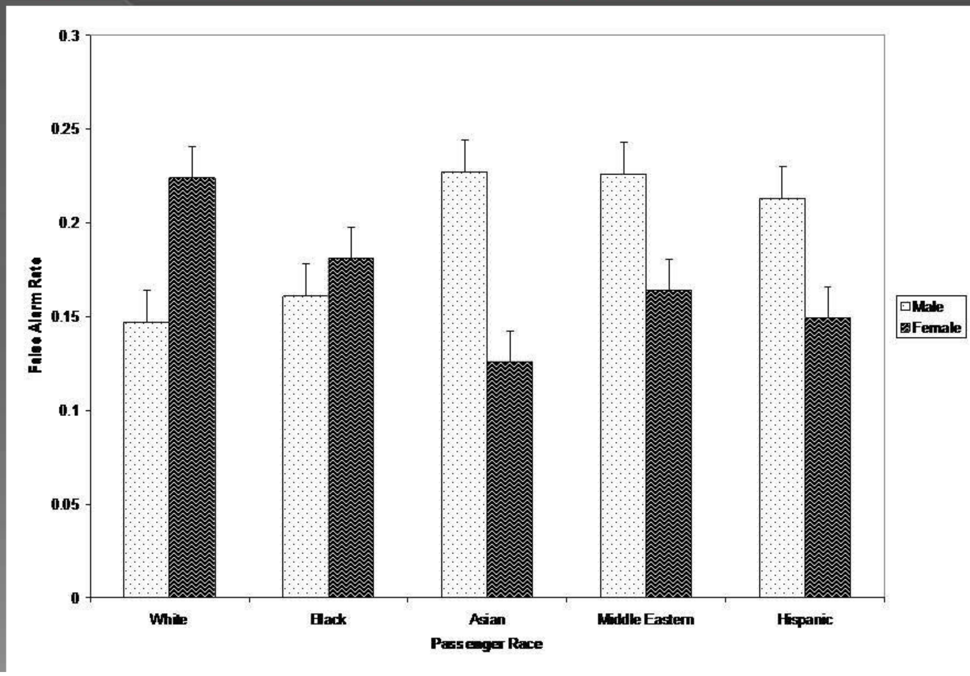
- **Pre-anchor** 2 (gender) X 5 (race) mixed measures ANOVA
  - passenger gender
    - $F(1, 23) = 5.131$, $p = .033$, $\eta 2 = .182$
  - passenger gender by passenger race interaction
    - $F(4, 92) = 3.120$, $p = .019$, $\eta 2 = .119$



False Alarm Rate: Low Time Pressure interaction

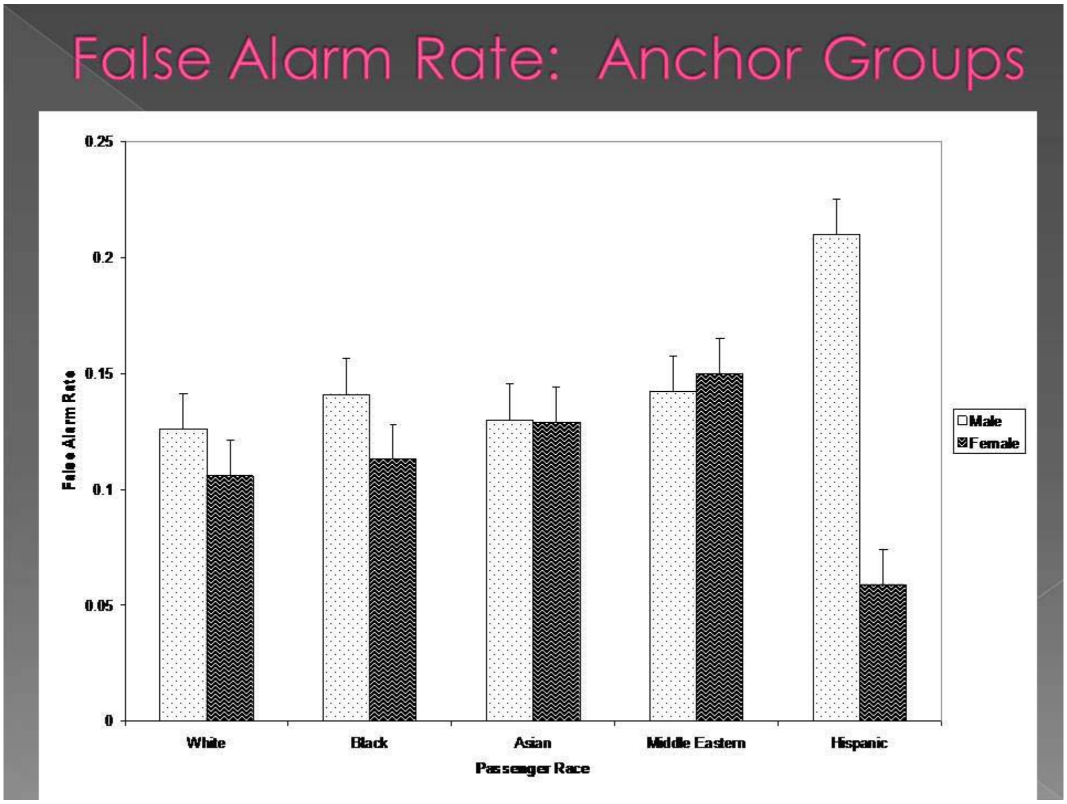
- No-anchor
  - passenger gender by passenger race interaction
    - $F(4, 92) = 3.221$, $p = .016$, $\eta 2 = .12$
- Post-Anchor
  - No significant main effects or interactions

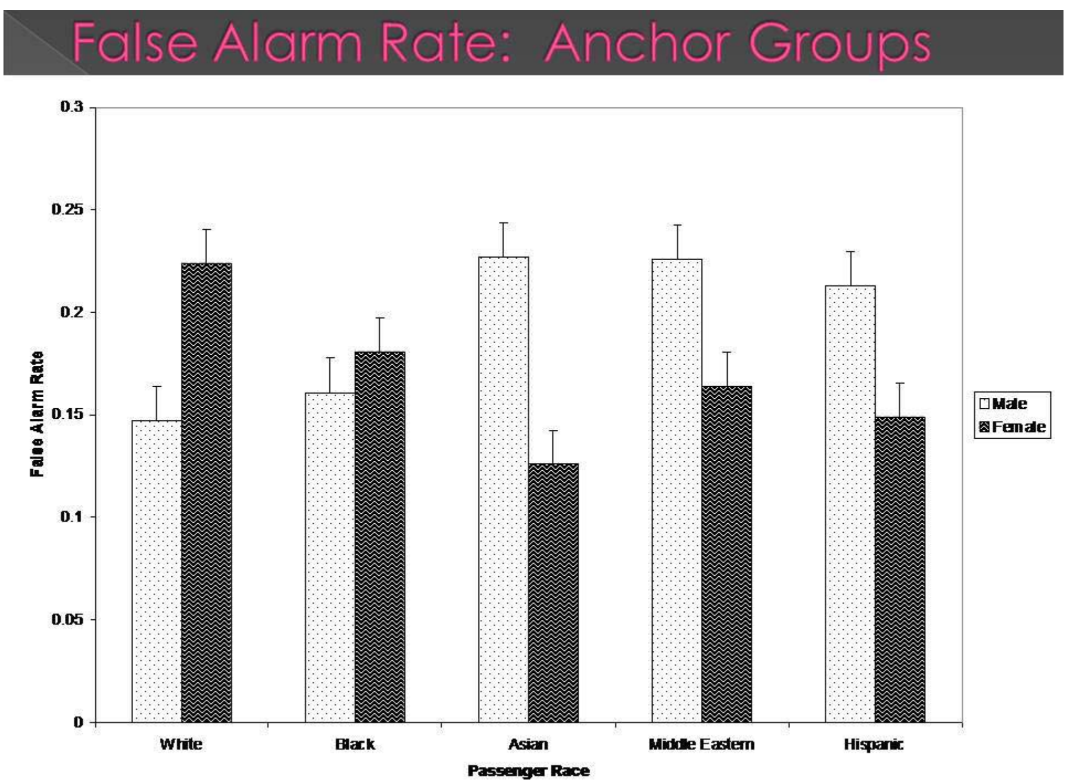
False Alarm Rate: Low Time Pressure interaction



False Alarm Rate: Anchor Groups

- 2 (gender) X 5 (race) X 2 (time pressure) mixed measures ANOVA
- Pre-Anchor
  - passenger gender by passenger race
    - $F(4, 88) = 3.132$, $p = .019$, $\eta 2 = .125$
- No-Anchor
  - time pressure
    - $F(1, 22) = 6.958$, $p = .015$, $\eta 2 = .24$
  - passenger gender by passenger race
    - $F(4, 88) = 3.145$, $p = .018$, $\eta 2 = .125$
- Post-Anchor
  - No significant main effects or interactions


False Alarm Rate: Anchor Groups
0.25
0.2
0.15
0.1
0.05
0
False Alarm Rate
White
Black
Asian
Middle Eastern
Hispanic
Passenger Race
Male
Female


False Alarm Rate: Anchor Groups
0.3
0.25
0.2
0.15
0.1
0.05
0
False Alarm Rate
White
Black
Asian
Middle Eastern
Hispanic
Passenger Race
Male
Female

Sensitivity (d')

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA
- time pressure
  - $F(1, 66) = 47.34, p \leq .001, \eta2 = .418$
- passenger gender, passenger race, and anchor
  - $F(8, 264) = 3.34, p = .001, \eta2 = .092$

Sensitivity (d')

- Non-significant main effects and interactions

Source				
Main Effects	**df**	***F***	**Sig**	**η^2**
Passenger Gender	1, 66	2.99	0.088	0.043
Passenger Race	4, 264	0.707	0.59	0.011
Anchor	2, 66	1.64	0.2	0.047
Interactions	**df**	***F***	**Sig**	**η^2**
Passenger Gender X Time pressure	1, 66	0.329	0.57	0.005
Passenger Gender X Anchor	2, 66	0.32	0.73	0.01
Passenger Gender X Time Pressure X Anchor	2, 66	0.242	0.79	0.007
Passenger Race X Time Pressure	4, 264	0.283	0.89	0.004
Passenger Race X Anchor	8, 264	0.765	0.634	0.023
Passenger Race X Time Pressure X Anchor	4, 264	0.601	0.78	0.018
Passenger Gender X Passenger Race	4, 264	1.22	0.3	0.018
Passenger Gender X Passenger Race X Time Pressure	4, 264	1.65	0.16	0.024
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	1.4	0.2	0.041

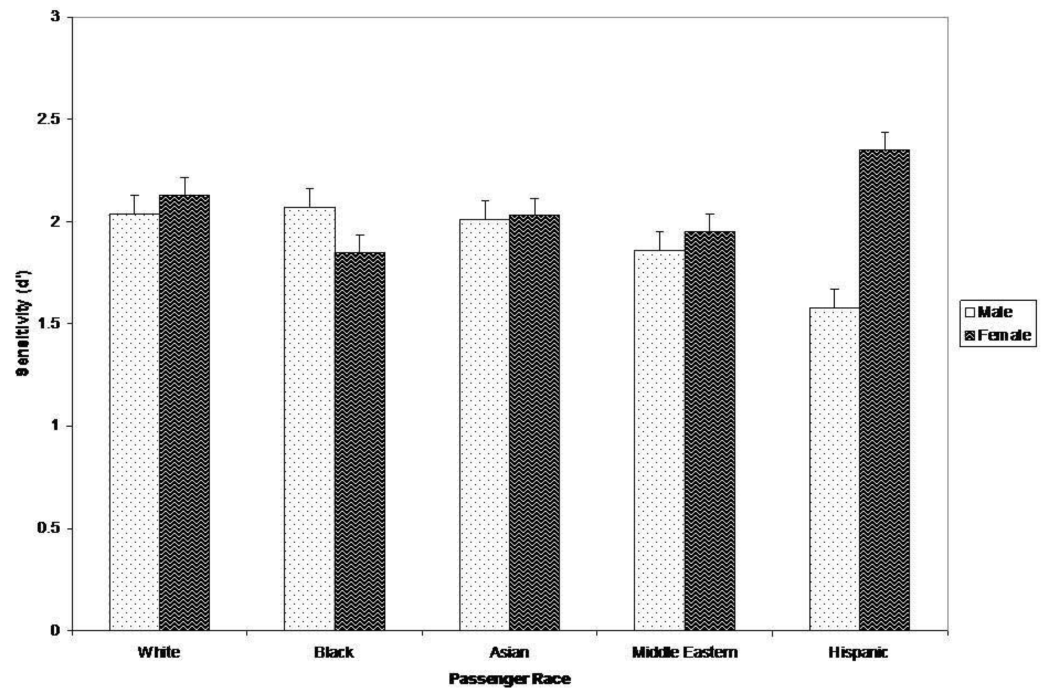
Sensitivity (d'): Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
- High Time Pressure
  - passenger gender bypassenger race by anchor
    - $F(8, 132) = 2.144$, $p = .036$, $\eta 2 = .115$
- Low Time Pressure
  - passenger gender by passenger race by anchor
    - $F(8, 132) = 2.607$, $p = .011$, $\eta 2 = .136$

Sensitivity (d'): Anchor Groups

- 2 (gender) X 5 (race) X 2 (time pressure) mixed measures ANOVA
- Pre-Anchor
  - time pressure
    - $F(1, 22) = 24.068$, $p \leq .001$, $\eta 2 = .522$
  - passenger gender by passenger race
    - $F(4, 88) = 2.963$, $p = .024$, $\eta 2 = .119$

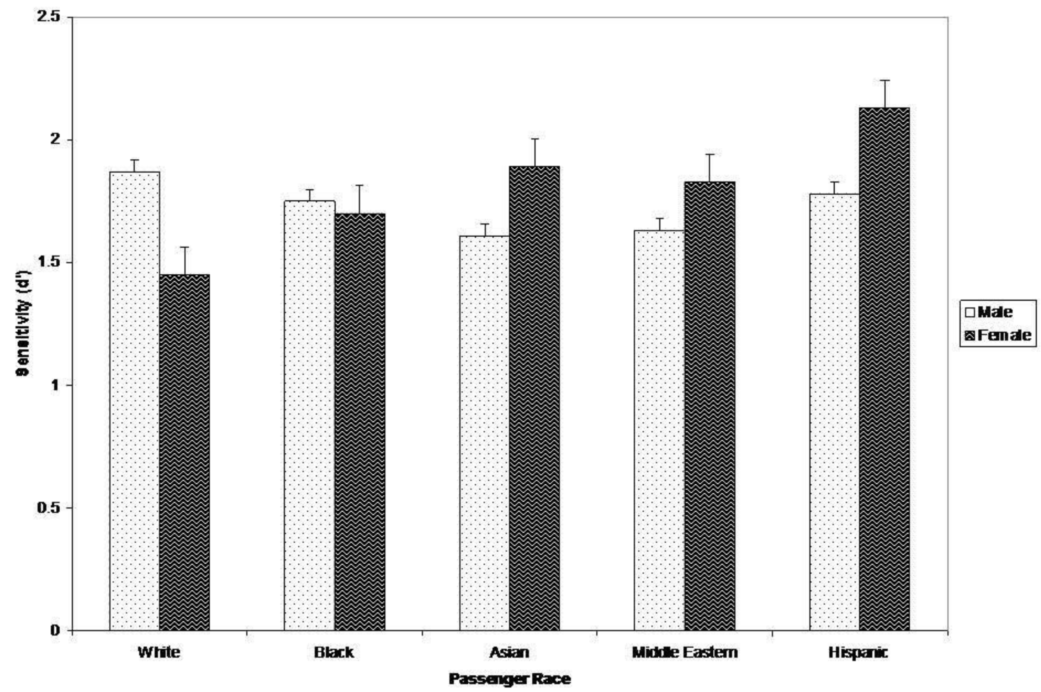
Sensitivity (d'): Anchor Groups



Sensitivity (d'): Anchor Groups

- No-anchor
  - time pressure
    - $F(1, 22) = 27.139$, $p \leq .001$, $\eta 2 = .458$
  - passenger gender by passenger race
    - $F(4, 88) = 2.56$, $p = .04$, $\eta 2 = .104$

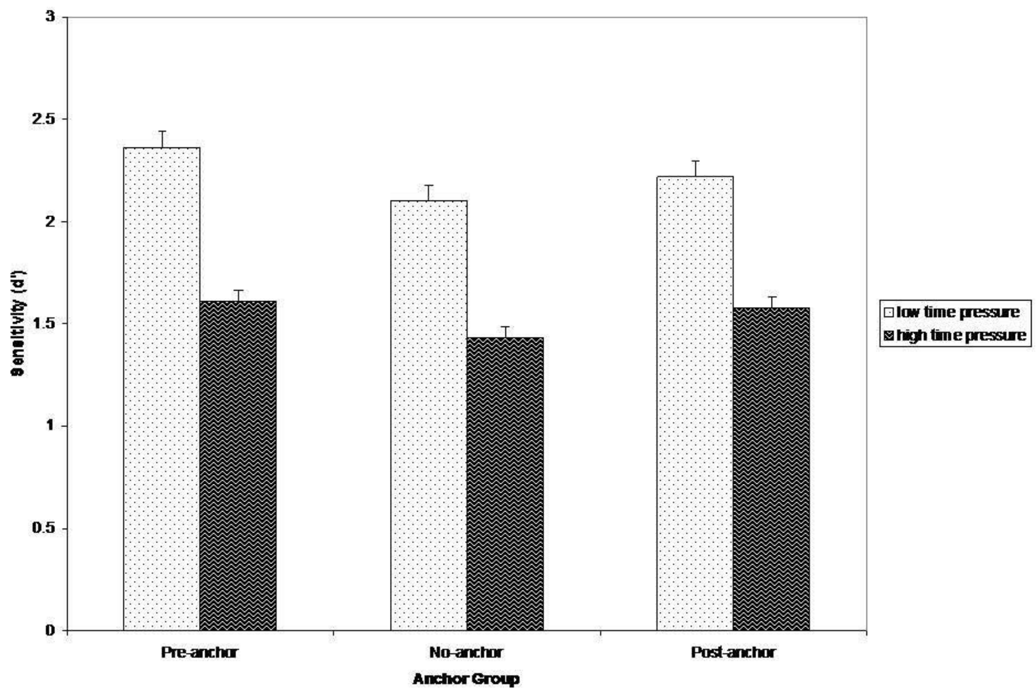
Sensitivity (d'): Anchor Groups



Sensitivity (d'): Anchor Groups

- Post-anchor
  - time pressure
    - $F(1, 22) = 9.717$, $p = .005$, $\eta 2 = .306$

Sensitivity (d'): Anchor Groups



Response Criterion Setting (c)

- 2 (time pressure: high vs. low) X 3 (anchor: pre-anchor, post-anchor, no-anchor) X 5 (passenger race: White, Black, Asian, Middle Eastern, Hispanic) X 2 (passenger gender: male vs. female) mixed measures ANOVA
- passenger race
  - $F(4, 264) = 8.48$, $p \leq .001$, $\eta 2 = .114$
- passenger gender, time pressure, and anchor
  - $F(2, 66) = 3.50$, $p = .036$, $\eta 2 = .096$

Response Criterion Setting (c)

- Non-significant main effects and

Source				
Main Effects	**df**	***F***	**Sig**	**η^2**
Passenger Gender	1, 66	1.8	0.184	0.027
Anchor	2, 66	0.784	0.46	0.023
Interactions	**df**	***F***	**Sig**	**η^2**
Passenger Gender X Time pressure	1, 66	3.45	0.068	0.05
Passenger Gender X Anchor	2, 66	0.044	0.96	0.001
Passenger Race X Time Pressure	4, 264	1.33	0.26	0.02
Passenger Race X Anchor	8, 264	0.378	0.93	0.011
Passenger Race X Time Pressure X Anchor	4, 264	0.584	0.79	0.017
Passenger Gender X Passenger Race	4, 264	0.431	0.79	0.006
Passenger Gender X Passenger Race X Time Pressure	4, 264	1.92	0.11	0.028
Passenger Gender X Passenger Race X Anchor	8, 264	1.78	0.08	0.051
Passenger Gender X Passenger Race X Time Pressure X Anchor	8, 264	1.11	0.36	0.032

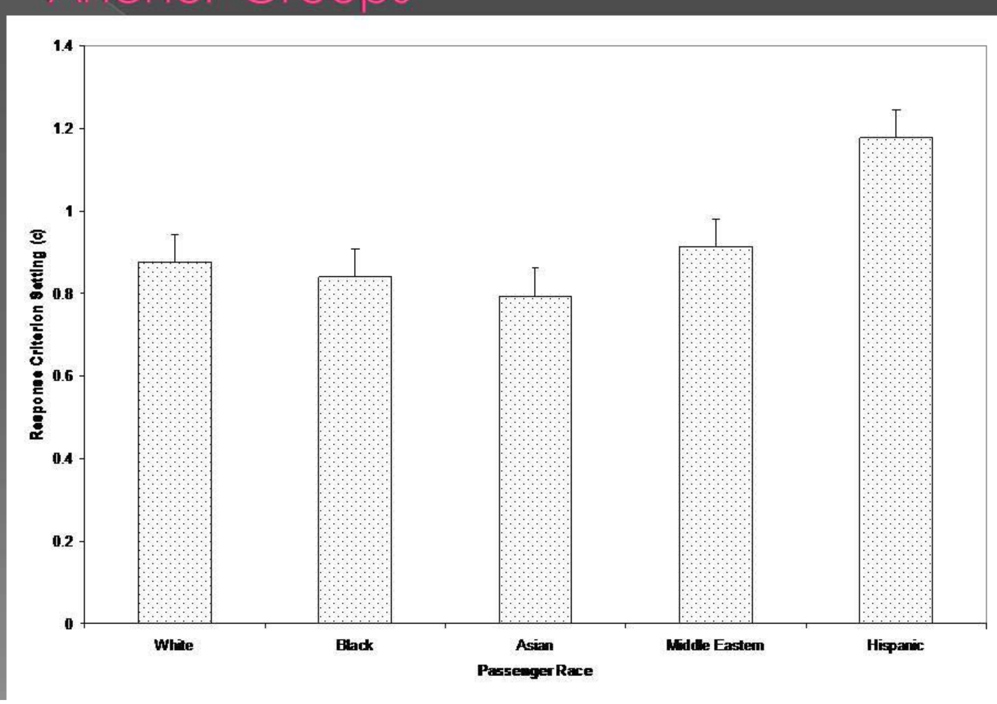
Response Criterion Setting (c): Time Pressure Groups

- 2 (gender) X 5 (race) X 3 (anchor) mixed measures ANOVA
- High Time Pressure
  - passenger gender
    - $F(1, 33) = 6.41$, $p = .016$, $\eta2 = .163$
  - passenger race
    - $F(4, 132) = 2.46$, $p = .048$, $\eta2 = .069$
  - anchor
    - $F(2, 33) = 3.523$, $p = .041$, $\eta2 = .176$
- Low Time Pressure
  - passenger race
    - $F(4, 132) = 6.56$, $p \leq .001$, $\eta2 = .166$

Response Criterion Setting (c): Anchor Groups

- 2 (gender) X 5 (race) X 2 (time pressure) mixed measures ANOVA
- Pre-anchor
  - passenger race
    - $F(4, 88) = 2.837$, $p = .029$, $\eta 2 = .114$

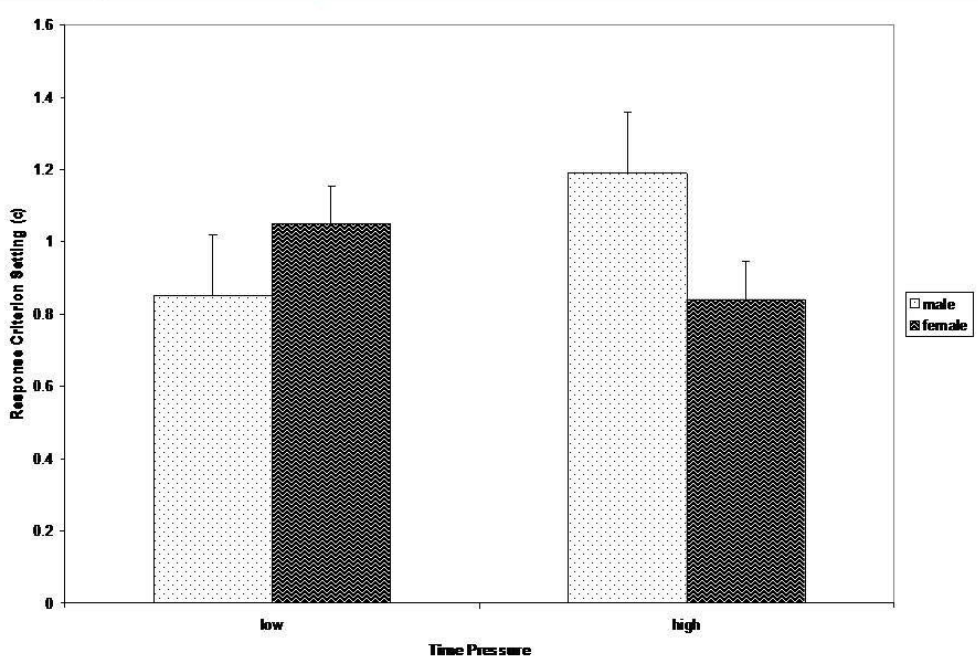
Response Criterion Setting (c): Anchor Groups

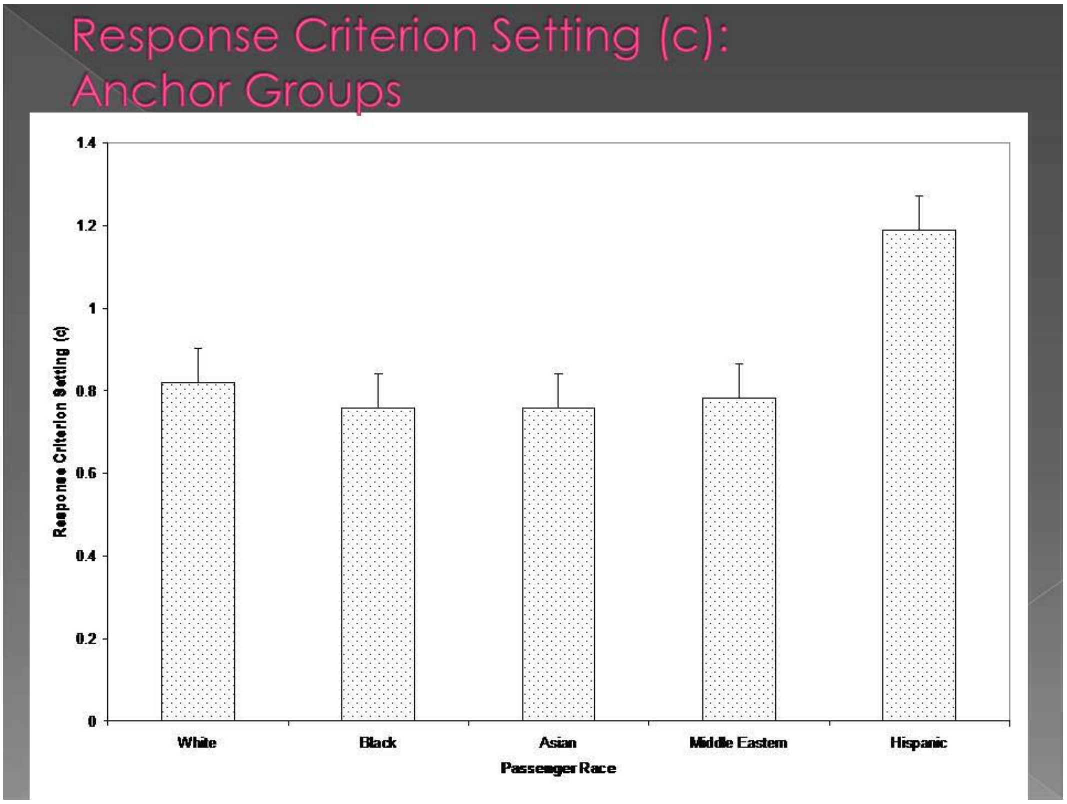


Response Criterion Setting (c): Anchor Groups

- No-anchor
  - passenger gender by time pressure
    - $F(1, 22) = 14.054$, $p = .001$, $\eta 2 = .390$
  - passenger gender by passenger race by time pressure
    - $F(4, 88) = 14.054$, $p = .001$, $\eta 2 = .390$).

Response Criterion Setting (c): Anchor Groups




Response Criterion Setting (c):
Anchor Groups
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Response Criterion Setting (c)
White
Black
Asian
Middle Eastern
Hispanic
Passenger Race


Discussion
Role of time pressure
Role of anchoring

Role of Time Pressure

- High time pressure
  - Fewer hits
  - Higher false alarms
  - Less time, can localize but not identify[2]
  - Higher response threshold for males
- Low time pressure
  - Lower response threshold for male passengers
    - Lower response criterion setting

Role of Anchoring

- Pre- or Post-anchor
  - Suppress social-cognitive biases
    - More salient when time pressure was low
- Decision heuristics
  - Gender
  - Race
    - Racial bias: higher false alarm rates for hispanics

Implications for Training and Design

- Luggage screening stations
  - Increase number of screeners
  - Block luggage screeners view of passengers
- Social-Cognitive Biases
  - Can be mitigated
  - Create training programs

Limitations and Conclusions

- Limitations
  - Participants
  - Simulation vs. real world screening
  - Tangible consequences
- Conclusions
  - Effects of social-cognitive biases
  - Design and training


Future Questions Contact email:
jrbrown@odu.edu