In: New Developments in Surveillance Systems
Editors: Katrina F. Hodges
ISBN: 978-1-63482-464-4

Chapter 3

COMPUTER SYSTEMS TO HEIGHTEN THE EFFECTIVENESS OF REAL-TIME SURVEILLANCE

Judith Gelernter
School of Computer Science,
Carnegie Mellon University, Pittsburgh, PA, US

ABSTRACT

Surveillance errors might occur because of fluctuations in attention, as when a person looks directly at a target without apparently seeing it, and without responding. When less attention is given to an infrequently-seen target, the brain might register what it *expects* based on the past rather than what is experienced actually. We heighten attention to infrequently seen targets by drawing on what is called the "prevalence effect", and repeating artificial instances of those targets. This chapter describes how we modeled the prevalence effect in an experimental system. Experimental findings demonstrate the method's potential to improve target threat detection during real-time surveillance by heightening attention.

I. INTRODUCTION

Target detection error. When the level of distraction is high because attention is absorbed elsewhere, or because what we are looking for is so rare that we do not expect it, we might look directly at something but not respond. This has been called *inattentional blindness* (Mack and Rock, 1998), or *inattentional deafness* (Dehais et al., 2013). Inattentional blindness and deafness are common in daily life owing in part to the narrow focus of human attention.

The controversy: attention error or vision error? Scientists dispute whether inattentional blindness is due to lack of seeing or hearing an event (Rees et al., 1999; Dehais et al., 2013), or whether it is due to seeing or hearing the event so briefly that it could not be stored in memory, and so was unable to be reported (Wolfe 1999).

Neurological evidence demonstrates that the error stems from lack of attention rather than from poor vision. It has been shown, for example, that objects to which we are inattentionally blind nonetheless register brain activity (Bressan and Pizzighello 2008), and that this brain activity is in the prefrontal cortex (Thakral 2011). Thus, the explanation for the inattentional blindness is that the brain sensed the events, but the events were not experienced long enough or enough times to be stored in memory, and so it seems as though they were unseen.

What do we mean by "target"? We define target as something being looked for or listened to (Steelman et al., 2011; Dehais et al., 2013). Target detection could focus on a knife buried in luggage, or a plane collision on an airport traffic controller screen (Wickens et al., 2009), or whatever obstructs a person's path while walking down a street (Dehaene and Changeux, 2005). Target events have some duration, as the behavior of visitors to a building, residents in a city, or crowds during a public event.

How common are misses in target detection? Various studies show that miss rates can be high, although the sample sizes and methods for each study differ. The gorilla-in-the-basketball-court study, where onlookers' attention is absorbed by basketball players to the extent that they do not detect a person who enters the court wearing a gorilla suit (Simons and Chabris, 1999) showed miss rates of 42% and 50% for repeated trails. A 2002 national evaluation of airport screeners showed a 25% miss rate, as did a 2004 evaluation at the Newark, NJ airport (Hallinan, 2009). Several studies suggested that the miss rate for radiologists examining x-rays was about 30% (Hallinan, 2009). An inattentional deafness experiment had a miss rate of 39.3% (Dehais et al., 2013).

Solution proposed: A prevalence-based framework for improved target detection. This chapter describes a system to improve real-time target detection, and therefore surveillance effectiveness. The improvement in target detection stems from including artificial targets to raise prevalence, and it could also come from intelligent system feedback response. A similar method was proposed but not implemented for closed-circuit television/video (CCTV) (Niel et al., 2007).

Improving attention by target repetition (prevalence). It has been conjectured that events seen infrequently are not reported because they are not remembered (Wolfe et al., 2007). The more a certain type of event is seen–the more prevalent—the longer that event type should remain in memory because it becomes more expected. When a similar event is seen, the event will be looked at for a slightly longer time (Cisek et al., 2009) such that the person is better able to spot and report the recurrence.

The current chapter validates a method for improving attention that has been demonstrated previously in laboratory tests (Wolfe et al., 2007; Evans et al. 2011b; White and Davies 2008; Nakashima et al., 2013; Schwark et al., 2013). We implemented this method in a realistic framework to show that the more event–targets witnessed, the more likely it is that those event–targets will be able to be reported.

Improving attention via other approaches. Staal (2004) lists factors that tend to inhibit the ability to keep one's mind on something: stress, lack of sleep, alcohol, noise, or time pressure, for example. Conversely, a number of factors cited below have been shown to lessen inattentional blindness, and help people remember and report what was experienced.

Internal factors that might lessen inattentional blindness

- the person is directed to what he should perceive (Mack 2003)

- the main task is less onerous (Green 2004) or has fewer cognitive demands (Macdonald and Lavie 2008)
- the person is relaxed or under relatively little stress (Bishop et al. 2009)
- the person is motivated by incentives such as money (Pessoa 2009)

External factors that might lessen inattentional blindness

- external events are preceded by cues timed at regular intervals (MacLean et al. 2009)
- external events are more conspicuous (Green 2004), such as an object in motion when other objects are static (Klotz 2007)
- external events are emotionally evocative -- as inferred from a study of a short-term form of inattentional blindness called attentional blink (Asplund et al. 2010)
- external events share properties of the main activity (Folk et al. 1992; Most et al. 2007)
- external events are of different sensory stimuli than the main event (Wayand et al. 2005; Sinnett et al. 2006)

Most of the internal and external factors cited above cannot be modeled visually in as straightforward a way as can prevalence, which is why our visual system is modeled on prevalence.

Other framework solutions to improve target detection. Some targets can be detected automatically, and even in real time. The targets are of the nature of a door opening, as can be detected by the Ocularis IS surveillance video alert system.[1] These are not comparable to complex human behaviors that are detected in this study. Targets can also be detected after the fact by reviewing video footage. Time can be compressed in video to show only salient changes, as in systems like the video review analysis tool, Briefcam.[2]

Other attempts to improve anomaly detection in video graphics use color (von Bastian et al., 2010), three-dimensional representations (Megherbi et al. 2012), or image blending (von Bastian et al. 2010).

Our experiment. We tested in the domain of security because of the high number of mistakes reported in Transportation Security Administration evaluations (Mosk et al. 2010). We created a set of three videos showing everyday activities inside a building, and these activities are interspersed with threats (targets). The main independent variable is the number of targets, and the dependent variable is target detection accuracy.

Research questions considered are:

- Is it easier to detect threats that are seen less or more frequently?
- Can we model the threat-detection solution in terms of a threat frequency that is optimal?
- Does the proposed threat-detection solution result in user false alarms that make the solution infeasible?
- Does the proposed threat-detection solution introduce so much additional effort on the part of the user that simultaneous tasks are jeopardized?

[1] http://www.onssi.com/downloads/manuals/Ocularis_4.1/Ocularis_IS_41_spec_sheet.pdf.
[2] http://briefcam.com/.

- How is the proposed threat-detection solution influenced by the user's gender, age, and amount of sleep?
- How is the proposed threat-detection solution influenced by length of shift (amount of time the person has been working)?
- How would a threat-detection computer system or protocol look that implemented this solution?

Contributions

Quantitative data to answer the research questions was collected. Highlights among experiment results and contributions include:

Efficacy of prevalence method demonstrated in a security setting. It has been demonstrated that repeating target prevalence might improve detection (Wolfe et al. 2007). Our findings demonstrate that the prevalence effect does indeed lessen misses of rare events.

Science of the prevalence effect. We collected preliminary data toward answering the question of whether there exists an optimal artificial event frequency that will lessen inattentional blindness and improve target detection.

Creation of surveillance test videos to be released for further experiments. We intend to release our experimental security videos in three levels of threat prevalence to researchers for the purpose of advancing inattentional blindness science.[3]

Design of system to improve rare event detection. Our interest is to design systems based on the prevalence effect that will lessen the number of misses. The artificial event system created for these experiments serves as a prototype for actual systems which, with only minor changes, could be adapted to another security or medical domain.

II. Experiment

2.1. Creation of the Video Set for the Experiment

We modeled our security video for the experiment on video from actual security cameras that displays multiple interior views.[4] Typical of present-day security video is four or more color panels showing different views of the surveillance space, with the views changing every few minutes. In simulation of actual security footage, our video is multi-panel, ~~in~~ color, and has no sound (see Figure 1). Recording audio is not legal in the author's state of Pennsylvania, although recording images is legal if people are told or are given a sign that they are being recorded.

We shot the footage in hallways with no natural light so that we could remain for several hours without changes in daylight marking the passage of time, as this would mar the final 4-panel video composite. Our video had one frame rotating out each minute, so that one-quarter of the screen would change regularly and in order.

[3] Please write to the author at gelern@cs.cmu.edu.

[4] We consulted with a representative of the WorldEyeCam company http://www.worldeyecam. com/security-cameras/.

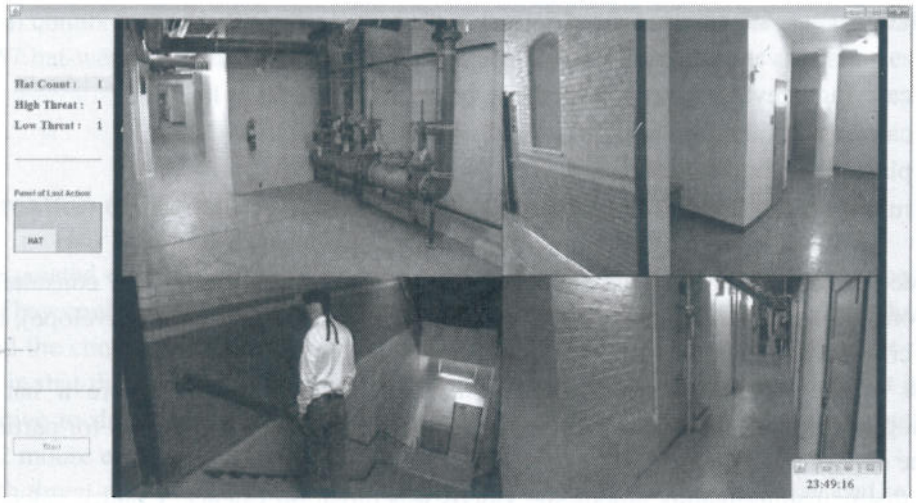

Figure 1. Screenshot of the 4-panel demo video, with counters at the far left for the user to see how he had responded to what he saw.

We hired a professional video company to record the action and assemble the video stream into the 4-panel display shown in Figure 1. The actors were mainly volunteer students from Carnegie Mellon University and its School of Drama, and some came from the community. We videotaped ordinary and threat activities from 10 angles around one of the older buildings on campus.

Main activity in the video. The videos showed everyday activities such as chatting and walking down the halls, sitting in hallway chairs, or waiting for the elevator.

Although the hallways were at times empty, "activities" included tying a shoelace, holding a bouquet of flowers, putting on lipstick, tearing an article out of a newspaper, giving performance tickets to a friend, and taking off a jacket while walking. The actors changed clothes at least once so that a casual onlooker would not recognize that the crowd was limited. Some of the actors wore hats; most wore regular clothes. The purpose of the hats was for the participants who would watch the video later to count and remember aspects of the hat and each hat wearer's behavior.

Threatening events (targets). Our threatening events were of two levels – low and high – as have been designated as suspicious by the National Terror Alert Center of the US Department of Homeland Security.[5] Some of the low level threat activities are ordinary such as sketching, but they become sinister with intent, such as sketching an area of a building with the purpose of infiltrating or destroying that area based on the sketch. The high level threats potentially result in immediate death or destruction.

Low threat activities were defined by the National Terror Alert Center as:

- sketching a building interior, or videotaping on a mobile device or looking with binoculars
- hiding from a surveillance camera,
- tampering with building fire or safety equipment,
- loitering alone for more than 5 minutes without apparent reason

[5] http://www.nationalterroralert.com/suspicious-activity/.

High threat activities were defined as:

- carrying a revealed weapon such as a pistol, rifle, sword, or knife
- carrying ammunition or bullets
- planting an explosive device
- running away from a suitcase (that might hold explosives)

To determine how the threats should be acted, a terrorism expert was consulted.[6] He advised on how to construct artificial terror devices (such as an exploding envelope), how to perform criminal gestures, and how the onlookers would likely behave.

Most of the action was just ordinary hallway activity. No one wore a hat while performing any threat, so as not to confound identification of hats and threats for participants taking the experiment later.

The distinguishing feature among the videos was threat prevalence. Control video A had only two threats. Video B had the two threats in A plus seven additional threats. Video C was the high prevalence case, and it included the two threats in A, the seven new threats in B, and with 16 additional threats, so that video C had 25 threats in all.

Most of the same appearances of hat-wearers in video A were also in B and C. A had 48 hat-wearer appearances, B had 57, and C had 63. Hat appearances and threat incidents were randomly distributed in time and location around the building, as well as in the quadrant in which they appeared on the screen monitor. Repetition among videos added experimental consistency.

Each threat activity within each video was unique. Unlike other attention-related studies, our threats differed in length, in obviousness, and whether one person or a group instigated the threatening activity. Threats that recurred in type were acted by different people and in different areas of the building. Thus, all events in our "threat library" were novel in action (although the same threat was at times acted by different people). This is important because repetition of the *exact* artificial events has been cited as a possible obstacle to the method's effectiveness (Schwaninger 2006), an obstacle which does not apply to our video conditions. Furthermore, during pilot testing, we removed differences in event conspicuousness, which has been found to influence inattentional blindness (Green 2004).

Video realism. The hallway activities were natural, and the threatening behaviors were as described by the National Terror Alert Center (see footnote 5). We included many more low level threats than high in an attempt at realism. Even so, numerous threats would probably not be seen within the space of a few hours.

Index to the videos created for scoring participant responses. Videos A, B and C were each two hours long, with much of the same hallway activity footage, occasional appearance of hat wearers, and suspicious threat activities so that participant responses among video groups could be compared consistently.

We manually created an index of event (hat or threat type), screen quadrant, and start and end time, to correspond to each video. For example, the index reflected the fact that video A had two threats, appearing at which time and in which quadrant each threat occurred.

[6] R. Conway, instructor of a course called "Terrorism and the Muslim Brotherhood". Communication with the instructor in April 2012.

Final counts for the index were that video A had 48 hat-wearer appearances and 2 threats; B had 57 hat-wearer appearances and 9 threats; and C had 63 hat-wearer appearances and 25 threats.

2.2. Creation of the Interactive Response System

We created *Call and Count* software so that users could record what they noticed in the video. They could "call the police" upon noticing a threat, and register counts for hats.[7] We designed the controls so that the keyboard input for hats required less manual effort than for threats, in that there were more hats than threats. We also designed the input so that it would be intuitive to differentiate between a response to a low threat (1 mouse click) and a high threat (2 mouse clicks).

Both threat-call and hat-count were specific to screen location, so the user had to specify in which panel the event happened. Threats could be entered in any panel by moving the mouse cursor and clicking in that panel. For hats, four "H" caps covered four keyboard keys (R,C,U,N) (Figure 2), to correspond spatially to the four panels of the video monitor. When a hat was spotted, the "H" key should have been tapped that corresponded to the panel where the hat was seen. We designed a control panel at screen left (Figure 1) so that the participant could see in which panel he had indicated the hat or threat, with the response tally for hats and threats visible throughout the study. If the participant realized he had made a mistake, he could use the "Remove" bar to undo his last input (Figure 2), although the option to remove an input remained for only a short time. The participant was not provided with the ability to go backward and rewind the video, to add to the effect of monitoring in real time.

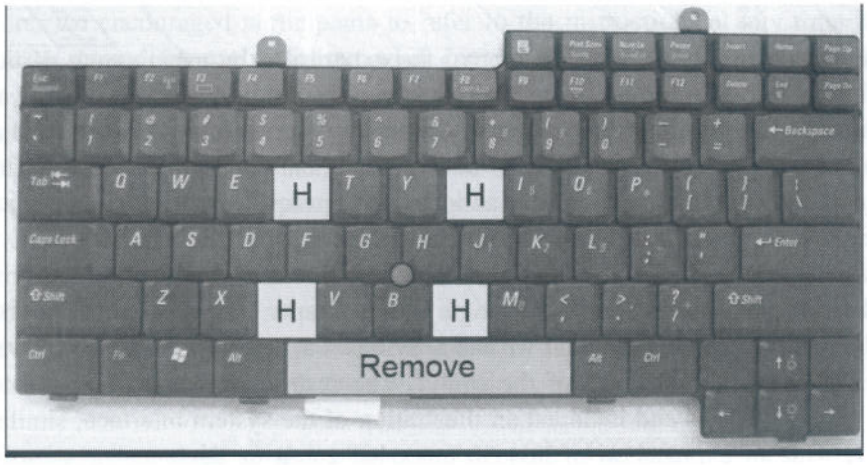

Figure 2. Keyboard showing input for hats by quadrant, and the "Remove" bar to remove the last action.

[7] The Java code relies on the Xuggle third-party open source software that plays the video, and Maven to manage the project, as well as a log file that dumps results directly into an Excel spreadsheet for end of study calculations. The video was recorded in the .wmv codec. With .wmv, each of the videos is about 2 GB large. They require either PC or Macintosh machines that are from about 2010 or later to run, with fast processors and good graphics cards.

2.3. Pilot Testing of Videos and Call and Count Interactivity Controls

We ran pilot experiments where we examined but did not collect participant data. Participant responses to the pilot experiment and their exit feedback afterwards showed what was effective and what was ineffective about the experiment, as well as aspects of validity. In this way, we examined the surveillance video, the Count and Call software, and the instructions procedure preceding the video monitoring.

We improved the experiment on the basis of the pilot. Our initial instructions session, for example, lasted 15 minutes. But we discovered during months of pilot testing that participants were not remembering what behaviors constituted threats. So we lengthened the pre-experiment instructions session to about 30 minutes, and we encouraged participants to refer to the definition of threats during the experiment. Also, we noticed which threats they missed more than others, and we adjusted the scoring for the experiment to remove this source of systematic error since it was impossible by that time to edit the videos further. Specifically, it was acceptable if one of these events was spotted, but the participant did not lose credit if the event was missed.

2.4. Running the Experiment

Population sampling. The population was recruited via the Carnegie Mellon University website for the Center for Behavioral and Decision Research.[8] We had 108 participants in three groups, with 36 per group.

Experimental design. Participants were divided randomly into three groups, with the groups differing according to number of threats per video (A, B, or C) in a between-group design.

Experiment testing procedure. We started the experiment by asking each participant to consider himself to be a building guard. Participants were not told that they were watching a video to add to the effect of real-time monitoring of surveillance camera footage. The instruction sheet did not mention the A, B or C groups, and participants were not told what the A, B and C signified when they were assigned to a group when they entered the room to begin the experiment. In that way, the experiment was set up without bias.

Instructions that described the tasks were handed to each participant. The instruction sheet consisted of the front and back of a single sheet of paper, describing a hat (a hood is not considered a hat, for example) and hat-wearer activities. Participants were asked to remember hat-wearer activities until the end of the study.[9] The instruction sheet described low- and high-level threat activities and included an illustration of the system interface, similar to that in Figure 1, with an explanation of how to enter hat count or threat call responses in the correct screen quadrant.

Participants began by reading the instructions to themselves. Then they were given the opportunity to ask questions in front of the group about what they had read, and all participants heard the answers. We then removed the instruction sheet from participants

[8] http://www.cbdr.cmu.edu/.

[9] This was for the purpose of increasing the cognitive load of the counting task (Cartwright-Finch and Lavie 2007), although we did not discuss the cognitive load aspect with participants.

temporarily, and asked them questions about the instructions. Next, we provided pages that were blank, and asked participants to list what they remembered they would be looking for with respect to hat-wearers and low and high threats. We went around the room examining each person's written responses and helping each individually to recall important details of the study. The purpose of this pre-experiment procedure was to reinforce participants' memory of low and high-level threats so that we would test *attention to* rather than identification of threats.

Setting for experiment testing. Our software was installed on Microsoft Windows machines in a college computer cluster. We reserved a room so that our experiment would be the only activity at that time. The instructions were read sitting near a computer, and then each participant began with the introductory questionnaire, demo video to practice the controls, and then the two-hour experiment.

Data from participants: Introductory questionnaire. Data from each participant remained anonymous. Using a short online form that preceded the experiment, we collected background on age and gender, whether the person had any prior security experience (although ultimately we did not have enough people with security experience register for the experiment to factor this in), and the number of hours of sleep the night before the experiment. These latter two factors were recorded because we thought they might affect performance.

Data not recorded: Demo video for practice. The 1.5 minute demo included one hat-wearer, one low-level threat, and one high-level threat, for the purpose of giving participants practice in recognizing hats and threats and responding using the controls. Each participant's score was shown automatically at the completion of the trial, and anyone uncomfortable with the procedure was encouraged to redo the demo.

Instruction sheets were returned to participants before each took the demo because we did not want to measure participants' ability to recognize what was and what was not a threat. Furthermore, we encouraged participants to refer to the instructions at any time during the experiment to remind themselves about what constituted threatening behavior if they were unsure.

Data from two hours of the experiment. Each participant watched and responded to the video of the assigned group for two hours. Counts were recorded when hats were spotted in ordinary activities, and mouse-click calls to the police were recorded for threats.

Data from post-task exit questionnaire. The first exit protocol included a series of questions about hat-wearers' locations and behaviors.[10] The second exit protocol inquired about participants' self-reported attention level at the beginning, middle, and end of the study.

Summary of data we recorded from participants

We recorded anonymously from groups A, B, C:

- Gender, age, and the number of hours of sleep the night before
- Each individual hat noticed, with time and screen quadrant
- Each individual threat noticed, with time and screen quadrant
- Total correct high-level threats
- Total correct low-level threats
- Total correct hats

[10] The purpose was to justify the study design of having participants memorize aspects of hat-wearer behavior because we wanted the hat count task to have a higher cognitive load.

- Hats missed
- Threats missed
- Threats missed in level
- False alarms: if hat or ordinary activity was identified as a threat mistakenly
- Exit questionnaire on perceived attention level

After each participant responded to (distraction) hats and (target) threats in the entire 2-hour video, the software calculated the total correct and incorrect responses on the basis of comparison to the index. Participant reaction time was not measured because of inherent differences in threat types and durations.

Some researchers have found that high prevalence can lower reaction time by seconds (Schwaninger, Hofer and Wetter, 2007); others have found that high prevalence can increase reaction time (Wolfe and van Wert 2010). While reaction time is of interest in visual search generally, in our domain of security surveillance, seconds lost by slower threat-spotters are minimal when compared to the inefficiency of false alarms and the danger of missing threats that are real.

III. Results of the Experiment

Here we repeat the research questions stated in this chapter's Introduction in order to organize the presentation of results. Count, percentage, and binomial data as in our results are not in general normally distributed, so non-parametric tests (e.g., rank-based tests) are appropriate. In our data, many of the results are charted as averages or medians in order to show central tendencies. We also include error bars to show the standard error of the mean or the dispersion of the data.

3.1. Research Question: Is It Easier to Detect Threats That Are Seen Less or More Frequently?

Our null hypothesis for this research question was that increasing the number of threats (in two levels) would neither increase nor decrease the number of threats detected as correct and partially correct, where partially correct is when a participant identified that an event was a threat but noted the level wrongly. This is the question that largely supports or discredits our method.

Each video had a different number of threats, so in order to equalize the number of errors among video groups, we calculated the error rate as a percentage. We added the total number of misses of low-level threats, the total number of misses of high-level threats, and the total number of threats that were found inaccurately at the wrong level, giving half credit to threats missed in level. This was divided by the total number of threats, and then multiplied by 100 to get the error rate. The trend shows the participant error rates decreasing as the number of threats climbs (Figure 3). The prevalence effect is apparent in that more threats were detected by group C than by group A.

The trend in detection of low and high threats improved from group A to B to C (Figure 3). Detection improved more for high threats than for low, and this is despite the fact that the groups found high threats harder to detect.

Across all groups, 73% of low level threats, and 66% of high level threats were detected. The trend suggests diminishing returns, so that further increase in the number of threats will result in less improvement (Figure 3). Groups B and C showed similar error rates (Wilcoxon rank sum test p=0.203), but there was a significant lowering of error between group A and groups B + C (Wilcoxon rank sum test, p=0.041).

These results support the alternative hypothesis that increasing the number of threats lowers inattentional blindness. Thus, the results support prevalence as an approach to improve the effectiveness of real-time surveillance.

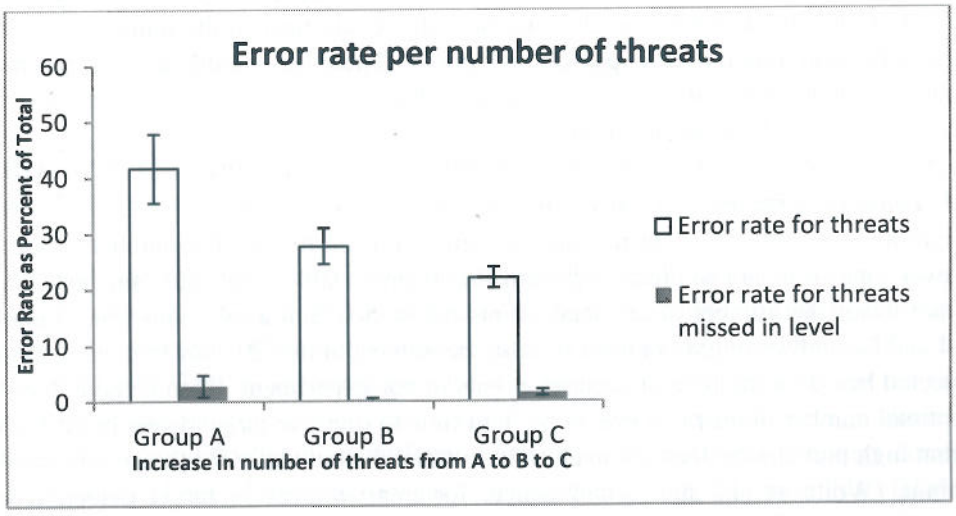

Figure 3. Chart bars show the average error rates by each 36-participant group, with the whiskers indicating the standard error of the mean. Group A had 2, B had 9, and C had 25 threats.

3.2. Research Question: Can We Model a Threat-Detection Solution in Terms of a Threat Frequency That Is Optimal?

Too few artificial events will not heighten user expectation enough to reduce misses, while too many artificial events will increase workload unnecessarily. It has been found that overly many artificial events may lead to an elevated false alarm rate (Schwark et al., 2013), and the slowing of response time, even when targets are absent (Wolfe and Van Wert, 2010). We did not find an increase in the false alarm rate, however (see Figure 6). We would like to know whether there exists an *optimal* number of artificial events per unit time to lessen misses of rare events in visual search tasks.

We need more data. Figures 4 and 5 show the per-group average error over the two hours of the experiment. The question is what trend the three points suggest. We experimented by trying to fit five different trend lines to our data: linear, exponential, logarithmic, polynomial, and power. We did not have enough data to determine which curve was correct, but we could

narrow the possibilities to the two best-fitting trend lines: the polynomial curve (Figure 4) and the power curve (Figure 5).

The Figure 4 polynomial trend line, should it best fit the data, suggests that we would get the maximum benefit in lessening inattentional blindness and minimizing errors if we randomly insert about 20 artificial events per two-hour period. But if the power trend line best fits the data as shown in Figure 5, then there will be no optimal number of artificial events per two hour period, and the usefulness of artificial event spotting will decay in accordance with the inverse power law. We cannot determine which of these trend lines fits best unless we test more participants in an additional threat condition – which would require making an additional video.

How could we get more data? We could create that fourth video for testing by editing out some threats from video C. This new video would need 20, 21, 22, or perhaps 23 threats (but not say, 24-26 threats, which is too close to the video C allotment); the number 20 is taken from the inflection point of the polynomial curve. Then we would test on another 36 participants. That additional data would allow us to determine whether the user response trend curve goes up (Figure 4) or down (Figure 5).

Event Frequency Window

In building an artificial event response system, we need to vary the number of artificial events over time so as not to create a pattern that is predictable. Our study suggests that we should not lower the number of artificial events below 5 within a two hour time frame (see Figure 4 and 5), and we might not need to raise the number above 20. Our frequency numbers are projected based on the type of security events in our experiment. Even if there does exist some optimal number of artificial events per unit time to improve target detection, it has been found that high prevalence does not need to be sustained continually for the cognitive effects to continue (Wolfe et al., 2007), and hence, for improvement in target detection to be maintained.

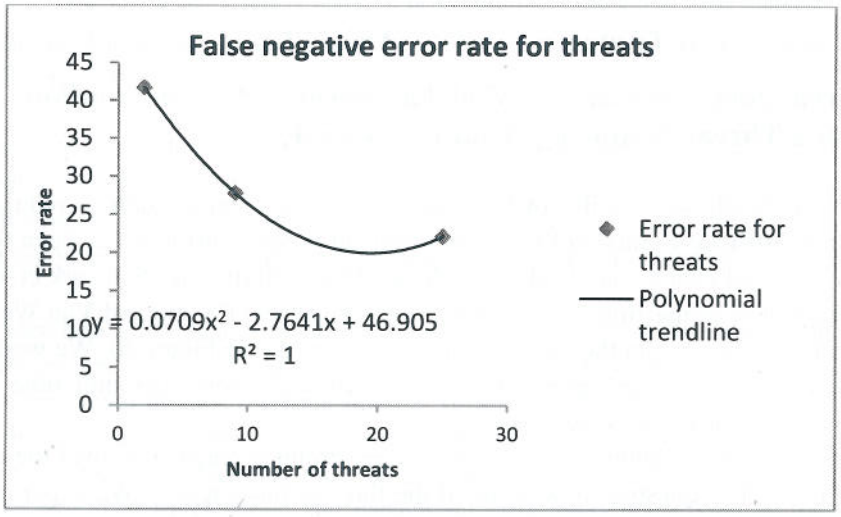

Figure 4. The average error rates for groups A, B and C, with the points fitted to a polynomial trend line.

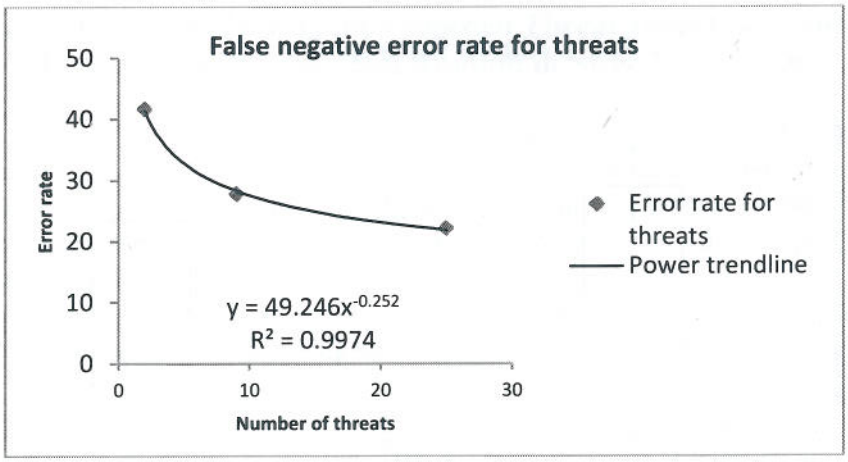

Figure 5. The average error rates for groups A, B and C, with the points fitted to a power trend line.

3.3. Research Question: Does the Proposed Threat-Detection Solution Result in User False Positives So As to Make the Solution Infeasible?

We tested the hypothesis that increasing the number of threats will neither increase nor decrease the number of false positives in threat detection. We found no significant increase in the number of incorrectly identified threats (Figure 6) (according to the Kruskal-Wallis rank sum test, p=0.160). This is important because it implies that our method would not bring on excessive false alarms from zealous searchers finding rare events when their awareness is heightened by a condition of high artificial event prevalence. Others' findings corroborate ours in that the false alarm rate in prevalence testing can be quite low (Hofer and Schwaninger 2005; Nakashima et al., 2013). Wolfe and van Wert (2010) found a marked increase in false alarms in a high prevalence condition, but this might have been caused by *overly* high prevalence.

3.4. Research Question: Does the Proposed Threat-Detection Solution Introduce So Much Additional Effort on the Part of The User That Simultaneous Tasks Are Jeopardized?

We tested whether increasing the frequency of threats affects accuracy at the hat counting task. The data show a small increase in the number of false negatives in the hat counting distraction task as the number of threats increases (Kruskal-Wallis rank sum test, p=0.0065). This very small p value indicates that the small increase in distraction task error did not happen by chance, but was probably a result of the increased requirements on the searchers. We chart the median rather than the mean in Figure 7 because the data have extreme scores, as shown by the whiskers in the plot.