DATA PAPER

Normed Images for X-ray Screening Vigilance Tasks

Stephanie Merritt[1], Lei Shirase[1] and Garett Foster[2]

[1] Department of Psychological Sciences, University of Missouri-St. Louis, US

[2] Maritz, US

Corresponding author: Stephanie Merritt (merritts@umsl.edu)

A great deal of interest concerns the study of vigilance performance, and trust in automation, due to their implications for public safety. This work provides an experimental resource for scholars in need of a vigilance style task. The dataset includes 150 X-ray images of luggage, and participants indicate whether or not they believe each image contains a dangerous item (simulating airport security screening). Using a sample of 991 adults recruited via MTurk, we normed these items in terms of difficulty. These stimuli can be used to study vigilance performance, trust in an automated decision aid, and other areas.

Keywords: Vigilance; Visual Search; Signal Detection; Automation

(1) Overview
Context

Collection Date(s)
Fall 2018

Background
Vigilance is defined as "the ability of organisms to maintain their focus of attention and to remain alert to stimuli over prolonged periods of time" [1, p. 433]. Vigilance tasks, then, are tasks that require sustained attention to identify rare, but important stimuli over a length of time [1, 2, 3]. Research suggests that humans are poorly suited to vigilance performance; a marked decrease in vigilance performance occurs after approximately 15-30 minutes [3, 4, 5, 6, 7]. Concern over this "vigilance decrement" and associated performance errors has inspired researchers to seek methods of alleviating the vigilance decrement [e.g., 8, 9, 10, 11, 12, 13, 14]. Other researchers have focused specifically on the vigilance task of monitoring automated systems for errors, and they have used vigilance tasks to examine automation use, complacency, and associated issues [e.g., 15, 16, 17, 18].

The X-ray screening task was popularized by Merritt and Ilgen [17], who used it to examine participants' trust in a fictional decision aid. They proposed that the task can be considered a "microworld" in that it contains essential elements of a real-life situation but also provides the opportunity for increased experimental control [19, p. 65]. Participants view a series of images of X-ray luggage containing various items, some of which may be dangerous (guns or knives). Participants' task is to indicate whether or not each image contains a weapon.

(2) Methods
Sample
991 U.S. adults (18+) were recruited via Amazon Mechanical Turk (MTurk) and compensated $4.00. Average age was 36 years (sd = 11.17). The sample was 56.6% male, 43.1% female, and 0.3% other genders. Racially, they were 73.7% white/Caucasian, 12.1% African-American, 4.6% Latin, 5.3% East or South Asian, 2.6% multi-racial, and 1.6% other ethnicities.

Materials
Scans of individual items (e.g., a baby bottle, a pair of shoes, a cell phone) were provided by the Transportation Security Administration for research in approximately 2003. The researchers then created luggage images by combining selected individual items into top-down views of scanned luggage. This was done by layering, rotating, and positioning the individual items as needed using Adobe Photoshop. This process resulted in 150 luggage images, 20% of which contained weapons. The images varied in the number of items contained and the degree the items overlapped. The images were consistent in terms of their overall size and luminance.

Procedures
To curtail respondent fatigue, each respondent viewed half of the 150 X-ray stimuli (each X-ray image was rated by ~500 respondents). For each image, if the participant

believed the image contained a weapon, they selected "search." If not, they selected "clear."

The study took participants 23.38 minutes on average. Each slide was viewable for as long as the participant liked, and there was no programmed delay between slides. In order to keep participants as naïve as possible, no performance feedback was provided. In addition to completing the X-ray task, participants completed 5 demographic items (before the X-ray task) and 81 self-report items related to another study (after the X-ray task).

Due to limitations of randomization in our survey platform, we created four sets of slide combinations. The combinations were equal in the number of slides containing weapons, and they ensured that each item appeared in two of the four sets and thus, should be evaluated by approximately N = 500 participants. Participants were randomly assigned to a slide set, and the order of the slides within each set was randomly assigned for each participant.

Quality Control
Respondents viewed a video-based training instructing them on what items to search for (i.e., guns and knives). The training video was just under 3 minutes long and is available in the repository. Following the training video, they were required to pass a multiple-choice attention check question about the instructions in the video; specifically they needed to correctly indicate that they should click "search" if the image contains a gun or knife, but not any other items. All respondents in our sample passed the attention check, with 92.8% passing on the first attempt and the remaining 7.2% passing on the second attempt.

Ethical issues
This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the University of Missouri-St. Louis. Informed consent was obtained from each participant. Data were identified only by a random ID number.

(3) Dataset description
Object name
X-RayImages-05-16-2019.zip
2019_09_19_Xray_Image_Norms.xlsx
XrayTraining.mp4

Data type
Primary Data
Processed Data

Format names and versions
The images are available as .jpg files. The slide norms are presented in Excel format, and the training video is mp4.

Data Collectors
The online data collection through MTurk was supervised by Dr. Stephanie M. Merritt in Fall 2018.

Language
English

License
CCby

Embargo
None

Repository location
https://irl.umsl.edu/psychology-faculty/61/. The files may also be found in the Harvard Dataverse Repository https://doi.org/10.7910/DVN/Z6R79K.

Publication date
2019/10/15

(4) Reuse potential
These materials can be used to produce new original research on vigilance, trust in automation, and more. Other relevant topics suggested by reviewers include sustained attention and fatigue, visual detection, decision certainty, threat biases, training, perceptual learning, and (if a per-image deadline is added) decision-making under stress. Researchers may also consider measuring per-slide response times as a component of these efforts.

The norms provided can be used to equate expected difficulty across different subsets of items. The average accuracy rate on the full set of slides was 70%. Researchers could use this information to set automation reliability at/above/below the level of the average participant. We suspect that our norms best represent the accuracy of inexperienced screeners.

We recommend that researchers create a scoring scheme for the task that matches the relative costliness of mistakes and the relative value of correct versus incorrect screening decisions. In this data collection, no formal scoring scheme was presented, and the average sensitivity was $d' = .24$ (average hit rate = 77%; average false alarm rate = 31%).

When reusing these images, researchers should consider whether to provide feedback on each individual decision, periodically, or not at all. Giving feedback more often is likely to produce greater improvements in performance throughout the task, compared to receiving feedback less often or not at all. Further, if working with an automated system, getting feedback affects the development of trust in the automated system.

Competing Interests
The authors have no competing interests to declare.

References
1. **Warm, J S, Parasuraman, R** and **Matthews, G** 2008 Vigilance requires hard mental work and is stressful. *Human Factors*, 50(3): 433–441. DOI: https://doi.org/10.1518/001872008X312152

2. **Davies, D R** and **Parasuraman, R** 1982 *The psychology of vigilance*. London: Academic Press.

3. **Parasuraman, R** 1986 Vigilance, monitoring, and search. In Boff, K R, Kaufman, L and Thomas, J P (Eds.), *Handbook of perception and human performance*, 2: 43.1–43.39. New York: Wiley.

4. **Helton, W S** and **Warm, J S** 2008 Signal salience and the mindlessness theory of vigilance. *Acta Psychologica*, 129(1), 18–25. DOI: https://doi.org/10.1016/j.actpsy.2008.04.002

5. **Nuechterlein, K H, Parasuraman, R** and **Jiang, Q** 1983 Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science*, 220(4594): 327–329. DOI: https://doi.org/10.1126/science.6836276

6. **See, J E, Howe, S R, Warm, J S** and **Dember, W N** 1995 Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, 117(2): 230–249. DOI: https://doi.org/10.1037/0033-2909.117.2.230

7. **Teichner, W H** 1974 The detection of a simple visual signal as a function of time of watch. *Human Factors*, 16(4): 339–352. DOI: https://doi.org/10.1177/001872087401600402

8. **Ariga, A** and **Lleras, A** 2011 Brief and rare mental "breaks" keep you focused: Deactivation and reactivation of task goals preempt vigilance decrements. *Cognition*, 118(3): 439–443. DOI: https://doi.org/10.1016/j.cognition.2010.12.007

9. **Correa, Á, Molina, E** and **Sanabria, D** 2014 Effects of chronotype and time of day on the vigilance decrement during simulated driving. *Accident Analysis & Prevention*, 67: 113–118. DOI: https://doi.org/10.1016/j.aap.2014.02.020

10. **Lara, T, Madrid, J A** and **Correa, Á** 2014 The vigilance decrement in executive function is attenuated when individual chronotypes perform at their optimal time of day. *PloS one*, 9(2): e88820. DOI: https://doi.org/10.1371/journal.pone.0088820

11. **Morgan, K, Johnson, A J** and **Miles, C** 2014 Chewing gum moderates the vigilance decrement. *British Journal of Psychology*, 105(2): 214–225. DOI: https://doi.org/10.1111/bjop.12025

12. **Ross, H A, Russell, P N** and **Helton, W S** 2014 Effects of breaks and goal switches on the vigilance decrement. *Experimental Brain Research*, 232(6): 1729–1737. DOI: https://doi.org/10.1007/s00221-014-3865-5

13. **Temple, J G, Warm, J S, Dember, W N, Jones, K S, LaGrange, C M** and **Matthews, G** 2000 The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, 42(2): 183–194. DOI: https://doi.org/10.1518/001872000779656480

14. **Young, M S, Robinson, S** and **Alberts, P** 2009 Students pay attention! Combating the vigilance decrement to improve learning during lectures. *Active Learning in Higher Education*, 10(1): 41–55. DOI: https://doi.org/10.1177/1469787408100194

15. **Merritt, S M** 2011 Affective processes in human–automation interactions. *Human Factors*, 53(4): 356–370. DOI: https://doi.org/10.1177/0018720811411912

16. **Merritt, S M, Heimbaugh, H, LaChapell, J** and **Lee, D** 2013 I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, 55(3): 520–534. DOI: https://doi.org/10.1177/0018720812465081

17. **Merritt, S M** and **Ilgen, D R** 2008 Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, 50(2): 194–210. DOI: https://doi.org/10.1518/001872008X288574

18. **Pop, V L, Shrewsbury, A** and **Durso, F T** 2015 Individual differences in the calibration of trust in automation. *Human Factors*, 57(4): 545–556. DOI: https://doi.org/10.1177/0018720814564422

19. **Lee, J D** and **See, K A** 2004 Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1): 50–80. DOI: https://doi.org/10.1518/hfes.46.1.50_30392