Experimental Brain Research (2018) 236:867–879
https://doi.org/10.1007/s00221-018-5184-8

RESEARCH ARTICLE

# Ceci n'est pas un walrus: lexical processing in vigilance performance

**Alexis R. Neigel[1] · Victoria L. Claypoole[2] · Gabriella M. Hancock[3] · Nicholas W. Fraulini[2] · James L. Szalma[2]**

Received: 28 September 2017 / Accepted: 17 January 2018 / Published online: 22 January 2018


**Abstract**

Vigilance, or the ability to sustain attention for extended periods of time, has traditionally been examined using a myriad of symbolic, cognitive, and sensory tasks. However, the current literature indicates a relative lack of empirical investigation on vigilance performance involving lexical processing. To address this gap in the literature, the present study examined the effect of stimulus meaning on vigilance performance (i.e., lure effects). A sample of 126 observers completed a 12-min lexical vigilance task in a research laboratory. Observers were randomly assigned to a standard task (targets and neutral events only) or a lure task (lures, targets, and neutral events presented), wherein lures were stimuli that were categorically similar to target stimuli. A novel analytical approach was utilized to examine the results; the lure groups were divided based on false alarm performance post hoc. Groups were further divided to demonstrate that the presence of lure stimuli significantly affects the decision-making criteria used to assess the performance of lexical vigilance tasks. We also discuss the effect of lure stimuli on measures related to signal detection theory (e.g., sensitivity and response bias).

**Keywords** Decision-making · Lexical processing · Lexical vigilance · Vigilance performance

## Introduction

Lexical processing relates to the semantic interpretation of signals (i.e., target stimuli) that include characters, symbols, words, shortened 'text message' type words (Head et al. 2012, 2013a), and in this vein, novel words (Head et al. 2013b). As a result, lexical processing is crucial for the successful performance of common, everyday tasks. For example, students assimilate reading materials and attend lectures as part of their education. Studying requires a great deal of focus, time, and attention, all of which are aspects of lexical processing.

The meaning individuals associate with a particular word, or similar abstract symbols, may influence lexical processing performance based on how the word or symbol is stored in their memory. Interestingly, little is known about the extent to which performance is affected by the meanings of target words or symbols. In one study, Yap and Seow (2014) demonstrated that positive and negative affective terms increased both response accuracy and response time in two lexical decision-making tasks. However, traditional decision-making indices of vigilance performance (e.g., sensitivity, response bias) were not calculated in this study, which makes it difficult to generalize these results to other lexical vigilance paradigms.

## Vigilance performance

Vigilance, or the ability to sustain attention for extended periods of time (Davies and Parasuraman 1982; Warm 1977), is routinely characterized by a decline in performance as function of time on task (i.e., the vigilance decrement; Mackworth 1948; Davies and Parasuraman 1982; Jerison 1970; Warm 1977). Historically, research has indicated that the decline in performance consistently observed in cognitive and sensory tasks is less reliable when utilizing vigilance tasks that rely on lexical processing (Deaton and Parasuraman 1993; Parasuraman and Mouloua 1987; cf; Majtásová and Šípoš 1974; Xiuying and Junying 1998).

✉ Alexis R. Neigel
alexis.neigel@gmail.com

1 Computational and Information Sciences Directorate, United States Army Research Laboratory, 321 Johnson Street, Aberdeen Proving Ground, MD, USA

2 Department of Psychology, University of Central Florida, Orlando, FL, USA

3 Department of Psychology, California State University, Long Beach, Long Beach, CA, USA

However, recent studies have demonstrated performance decrements in vigilance tasks using 'text-speak' (Head et al. 2012, 2013a), masked speech targets (Shaw et al. 2013), living versus non-living objects (Epling et al. 2016), and animals versus non-animals (Thomson et al. 2016; Thomson and Hasher 2017).

In the present study, we examine the effects of stimulus meaning on lexical vigilance performance using the task developed by Thomson et al. (2016). This vigilance task is unique given the inclusion of "lures", which are stimuli that are categorically similar to the target stimuli included in the task. Targets in this lexical vigilance task are words representing four-legged animals such as "cougar", "squirrel", and "llama". Lures, on the other hand, include words that describe animals that do not have four legs. The lures are, therefore, categorically similar, but clearly do not meet the criteria to be considered a critical signal. Some examples of these lures include "flamingo", "moth", and "ant". In this task, both targets and lures are included in a stimuli set consisting of many inanimate objects (i.e., "flag", "plate", "zipper", etc.). The target and lure terms appear with low frequency relative to neutral events. Observers are instructed to respond only to "four-legged creatures" throughout the vigil (Thomson et al. 2016, p. 76).

## Lexical vigilance and signal detection theory

Signal detection analyses rely on the assumption that a binary distinction exists between targets and non-targets (Macmillan and Creelman 2005). A vigilance task that incorporates lures effectively reshapes the decision-making criteria space, and consequently the underlying decision-making process, required to perform the lexical vigilance task (Fraulini et al. 2017). With the inclusion of lures, a third signal distribution is introduced, as observers must now distinguish between: targets, stimuli that are a priori determined to be non-targets but are conceptually similar to the target set, and distinct non-targets (see Fraulini et al. 2017, p. 527). The processing associated with traditional lexical vigilance tasks (see Epling et al. 2016; Head et al. 2012; Shaw et al. 2013) now require multiple stages of processing, which alters the meaning attached to the stimuli, as it relates to signal versus non-signal discrimination. For example, in the presence of lure stimuli, observers are presented with multiple response criteria: (1) the stimulus is a creature, and (2) the stimulus is specifically a four-legged creature (Thomson et al. 2016; cf; Fraulini et al. 2017). This addition complicates the decision process for observers. Rather than distinguish creatures from any other stimulus, they must: (1) identify the creatures, and (2) recall the numbers of legs. As such, a stimulus that is a critical signal in one task (say, a caterpillar), is a lure stimulus in another (due to its high number of legs).

One way to test the assumption that the additions of lures alter lexical vigilance performance is to examine performance based on the types of false alarms committed in the presence of lures. Although this is a novel approach, it is useful in that the effect of lure stimuli on observer lexical processing is relatively unknown in vigilance paradigms. For example, over the course of a vigil, differences in criterion setting and shifting should emerge between observers exposed to the lure stimuli, but to a lesser extent when observers are distinguishing between targets and non-targets. As vigilance tasks are by nature associated with a degree of criterion shifting, observers learn over the course of the vigil that the likelihood of target stimuli is relatively low (Baker 1959; Craig 1978, 1987; Deese 1955; Frankmann and Adams 1962; Fraulini et al. 2017; Vickers 1979). Thus, it would be useful to understand how the introduction of categorically similar non-targets (i.e., lures) affects sensitivity and response bias.

To summarize, the present study explores how decision-making changes when categorically similar stimuli are included in a lexical vigilance task, which facilitates shifts in perceptual sensitivity and response bias indices of vigilance performance. It is also possible that the presence of lure stimuli influences levels of perceived stress and workload that frequently accompany vigilance analyses.

## Method

### Observers

One hundred and twenty-six (75 female; 51 male) undergraduate students were recruited from the University of Central Florida research participation system. Observers completed this study for partial course extra credit. The average age of observers was 18.63 years (median 18.00 years, SD 1.29 years, range 18–26 years). All observers reported normal or corrected-to-normal vision. Observers were asked not to consume caffeine 24 h prior to participation in this study. All observers passed a test of English as a foreign language examination and/or spoke English as a primary language with fluency.

### Measures

#### Dundee State Stress Questionnaire

The Dundee Stress State Questionnaire (DSSQ; Matthews 2016; Matthews et al. 2002) was used to measure subjective levels of stress at pre- and post-task. The short version of the DSSQ consists of 20 items that measure three subsidiary factors: distress, task engagement, and worry (Matthews

2016). Higher scores indicate elevated distress, worry, or engagement in the task.

#### NASA-Task Load Index

The NASA-Task Load Index (NASA-TLX; Hart 2006; Hart and Staveland 1988) measures perceived workload across six dimensions including: mental demand, physical demand, temporal demand, performance, effort, and frustration. Observers provide a rating (0–100) for each scale, and then complete 15 pairwise comparisons. The ratings and pairwise comparisons are then used to calculate a weighted measure of global workload; higher scores indicate greater global workload.

### Experimental stimuli

Stimuli included the list of words used in Thomson et al. (2016), which was requested and obtained with permission from David R. Thomson via e-mail communication. Each word was programmed into SuperLab stimulus presentation software and was randomized for presentation via a desktop computer.

The Thomson et al. (2016, p. 76) protocol required observers to respond to "four-legged creatures" (i.e., targets) in a standard task that includes only distracters (i.e., non-animal stimuli such as "wire", "book", "slipper", etc.) or a task that includes a combination of distracters and lure stimuli (i.e., non-four-legged animals such as "bird", "snake", "ant", etc.). A full description of the word list and stimulus setup has previously been described elsewhere (see Thomson et al. 2016; Claypoole et al. 2017).

The standard task consisted of ten critical signals (i.e., four-legged animals such as "cougar", "llama", "squirrel", etc.) and 90 neutral trials (i.e., non-signals such as "apple," "phone," "wire," etc.) per block. The lure task consisted of ten critical signals, ten lures (i.e., non-four-legged animals; such as "canary", "lobster", and "walrus", etc.), and 80 neutral trials per block. This resulted in 100 trials per block across five periods on watch in each condition. Each word appeared for 200 ms, followed by a fixation cross that remained visible for 1100 ms, which resulted in a total response window of 1300 ms. The stimuli were programmed to appear in a random order and were programmed to ensure that lures and targets did not immediately follow one another.

### Apparatus and laboratory setup

Observers completed the study in a quiet, private laboratory with the facilitation of a research assistant. Stimuli were presented using SuperLab software (version 4.5) on a Dell OptiPlex 745 desktop computer. Observers were seated approximately 50.8 cm from the computer screen in a uniformly lit, private cubicle. Up to two observers could complete the experiment at a time, though it was most common to have only one participant complete the study at a time. When multiple observers were in the laboratory, each participant was instructed to complete the task individually. The research assistant was not present during the vigil.

### Procedure

Upon arrival, observers were randomly assigned to the standard task ($N = 67$) or lure task ($N = 59$). Observers were instructed to power down any electronic devices (e.g., cellphones, tablets, laptops, etc.) and surrender any timepieces (watches or otherwise) to the researcher prior to beginning the experiment. First, observers reviewed an electronic informed consent and asked the researcher any questions related to the research study. Informed consent was obtained from all the individual observers included in this study. Next, observers completed pre-task measures of stress.

The research assistant then prepared either the lure task or standard task in each research cubicle computer and asked observers to review the instructions prior to beginning the task. The instructional set used in the present study was identical to the task instructions provided in Thomson et al. (2016, p. 76). The research assistant indicated that observers should alert them when the task had ended by stepping outside of the research laboratory. The researcher then left the room at this time and waited in a nearby laboratory until observers completed the vigil, which was approximately 12 min in duration.

Upon conclusion of the vigil, the researcher returned to the room and requested the observers complete post-task measures of stress and workload, as well as a demographics form. Observers were then given a post-participation form detailing information about the study, and the appropriate participation credit was awarded.

### Data cleaning and outlier removal

To be included in the following analyses, observers needed to demonstrate at least a 70% correct detection rate in the first period on watch. Based on this performance criterion, seven observers from the lure condition and one observer from the standard condition were removed and excluded from analyses. Observers who committed excessive false alarms (in this study, excessive was defined as ten or more false alarms committed during the first period on watch) were also removed from the study. Consequently, an additional four observers from the lure condition were excluded from the analyses. This cleaning resulted in a usable sample of 126 observers.

## Results

The results indicated that two distinct groups of observers emerge when lure stimuli are included in lexical vigilance tasks: (1) a group of observers who commit false alarms to lure stimuli only ($N=28$), and (2) a group of observers who respond to a combination of lures and distracter stimuli ($N=31$). The mere presence of these two groups provides initial evidence that the addition of lures necessitates additional layers of lexical processing to respond appropriately to task stimuli.

### Stress analyses

A factorial analysis of variance (ANOVA) was performed between the conditions for each subscale of the DSSQ (see Fig. 1). There was a significant difference between the conditions across post-task distress scores, $F(2, 123)=4.97$, $p=.008$, $\eta_p^2=0.08$. Observers who committed both lure and distracter false alarms reported significantly greater post-task distress ($M=9.74$, SD 5.24) than observers in the standard task ($M=6.24$, SD 5.07, $p=.006$; note there were no significant differences in post-task distress between the lure conditions). There were no additional significant results to report for these analyses.

These measures were taken for comparison purposes as cognitive and sensory vigilance tasks typically produce a stress profile wherein task engagement decreases, distress increases, and no significant changes in worry are observed as a result of completing the vigil (Grier et al. 2003). The fact that the observed pattern of results does not correspond with the stress profile commonly seen in the vigilance literature is most likely reflective of the novel nature of the task. There is as yet no standard stress profile for lexical-based vigilance tasks, and this work represents the nascent steps in the establishment of such a touchstone.

### Workload analyses

A factorial ANOVA was performed between the conditions for global workload (see Fig. 2). There were no significant differences between the standard and lure conditions for global workload scores.

Following this, separate one-way ANOVAs with Bonferroni corrections were performed for all the raw NASA-TLX subscales to determine if significant differences were present between these factors. There was a significant difference between the conditions across frustration scores, $F(2, 123)=5.24$, $p=.007$. Specifically, observers who committed a combination of lure and distracter false alarms ($M=47.74$, SD 28.14) reported significantly more frustration with the task than observers who committed only lure false alarms ($M=28.68$, SD 29.32, $p=.031$) and observers in the standard task ($M=29.04$, SD 27.53, $p=.009$).

### Performance analyses

Separate mixed-measures factorial ANOVAs were performed for the following measures of performance using period on watch as the within-conditions factor and false alarm type [i.e., standard task (distracter false alarms are the

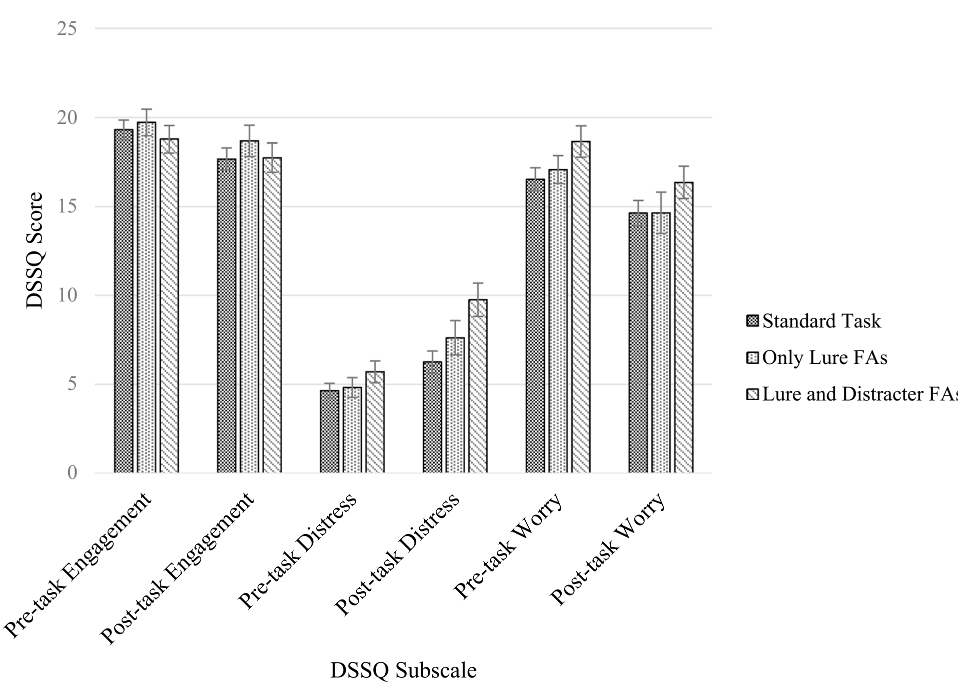



**Fig. 1** Subjective pre- and post-task stress scores across conditions. Error bars represent the standard error of the mean

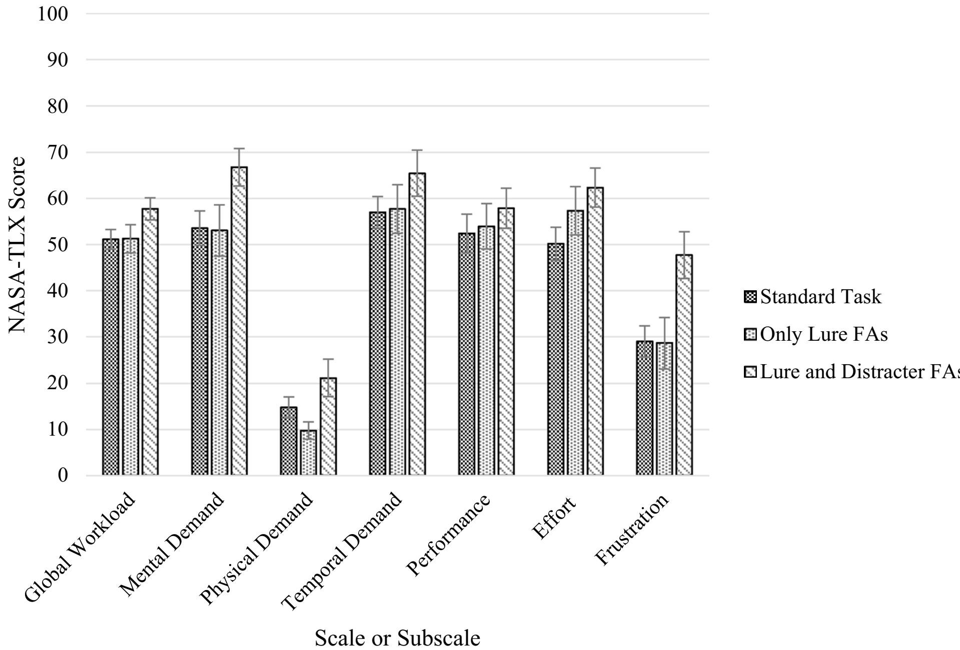



**Fig. 2** Subjective post-task workload scores across conditions. Error bars represent the standard error of the mean

only false alarms possible), lure only false alarms, lure and distracter false alarms] as the between-conditions factor to identify significant differences over time across proportion of correct detections, the proportion of false alarms, and average response time. Separate mixed-measures factorial ANOVAs were performed based on the length of the stimuli to examine the effects related to perceptual aspects of the stimuli. It should be noted that assumptions of sphericity were violated in some cases and a Huynh–Feldt epsilon statistic is reported accordingly.

### Correct detection performance

The proportion of correct detections (i.e., correctly identified target stimuli) observed over time is reported in Fig. 3. There was a significant main effect of period on watch, $F$(4,

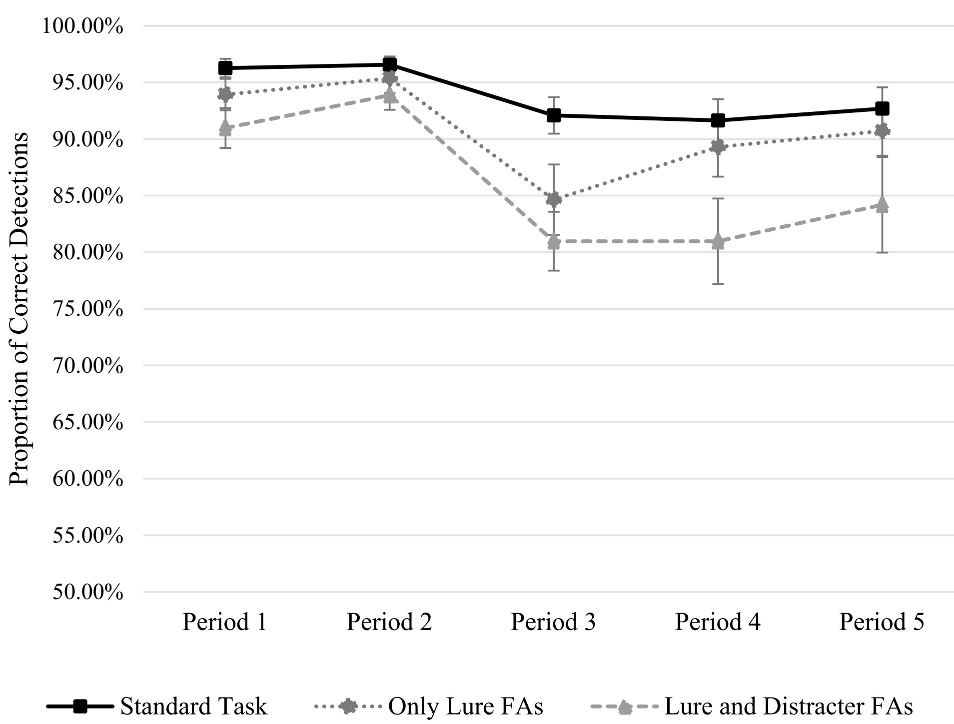



**Fig. 3** Proportion of correct detections as a function of period on watch and across false alarm groups. Error bars represent the standard error of the mean

120) = 16.32, $p < .001$, $\eta_p^2 = 0.12$, $\varepsilon = 0.73$. A decline in correct detection performance was observed between the first three periods of watch, whereas correct detection performance improved after Period three for all false alarm groups. The lure only false alarm group demonstrated performance that was closer to the distracter only false alarm group across Periods four and five. This observed pattern is not necessarily indicative of a traditional vigilance decrement characterized by a consistent decline in correct detections over time, though it should be noted that a vigilance decrement has previously been observed in experimental studies using this task (Claypoole et al. 2017; Thomson et al. 2016). The observed pattern in this study underscores a previous point: a decline in performance utilizing lexical vigilance tasks is not as reliably observed when compared to cognitive or sensory vigilance tasks.

There was a significant main effect of false alarm type, $F(2, 123) = 7.21$, $p < .001$, $\eta_p^2 = 0.11$. Pairwise comparisons indicated that the distracter only false alarm group committed significantly more correct detections than observers who committed a combination of lure and distracter false alarms ($p < .001$). However, performance was not significantly different between the lure only false alarm group and the lure and distracter false alarm combination group, or the distracter only false alarm group.

### Target length effects

In the Thomson et al. (2016) task, the stimuli are not balanced across letter lengths, which could lead to a shift in decision-making strategy unrelated to lexical processing (i.e., sensory-based processing). For example, observers may learn to perceptually identify a target or lure without fully reading the word (i.e., identify by perceptual aspects such as letter shape or letter combinations versus semantic content). Thus, we have further analyzed correct detection performance to explore the potential for word length effects.

Target word length effects are illustrated in Fig. 4. The average target word length is 5.78 letters in length (range 3–9 letters). Separate one-way ANOVAs with a Bonferroni correction for multiple comparisons were performed for all possible target letter lengths (i.e., 3–9 letters total in length). There was a significant difference between the groups for four letter target words, $F(2, 123) = 7.02$, $p = .001$, six letter target words, $F(2, 123) = 7.19$, $p = .001$, and eight letter target words $F(2, 123) = 5.83$, $p = .004$.

Post hoc tests indicated that observers in the standard condition ($M = 8.21$, SD 0.99) responded to more four letter target words than observers who responded to a combination of lure and distracter false alarms ($M = 7.22$, SD 1.36), $p = .001$. Similarly, observers in the standard condition ($M = 8.34$, SD 0.95) responded to more six letter target words than observers who committed a combination of lure and distracter false alarms ($M = 7.45$, SD 1.31), $p = .001$. Again, observers in the standard condition ($M = 3.88$, SD 0.33) responded to more eight letter target words than observers who committed a combination of lure and distracter false alarms, ($M = 3.52$, SD 0.72), $p = .007$.

There were no additional significant results to report for these analyses, though it should be noted that there were several commonly missed targets across the groups.

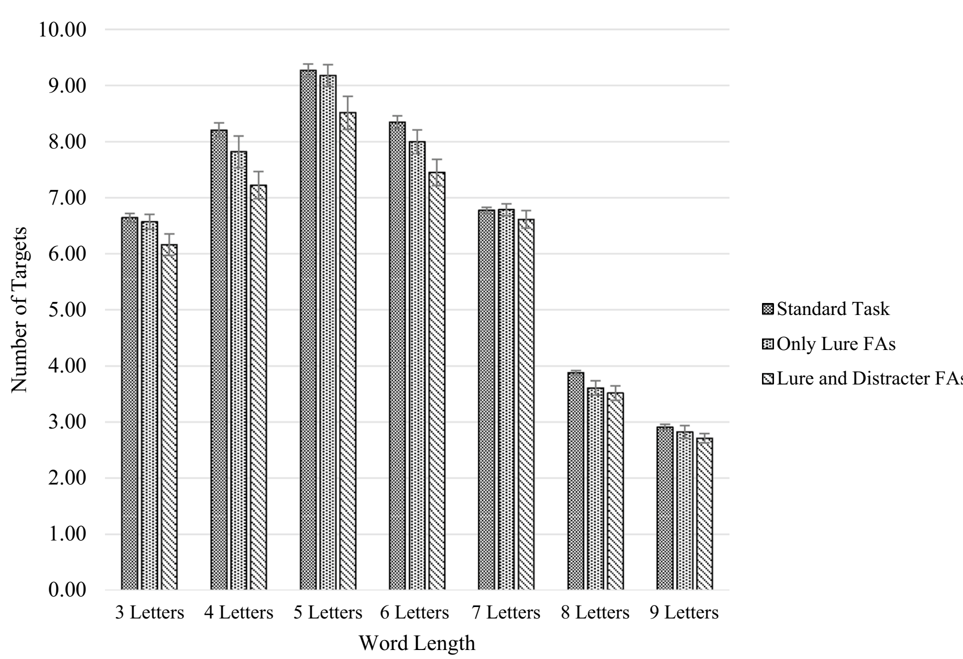



**Fig. 4** Average number of targets committed by false alarm group. Error bars represent the standard error of the mean. Note there are 7 three letter target words, 9 four letter target words, 10 five letter target words, 9 six letter target words, 7 seven letter target words, 4 eight letter target words, 3 nine letter target words, which total the 50 words included in the target stimulus list

These included “Bison” (15.87% missed), “Frog” (23.01% missed), “Gerbil” (22.66% missed), “Gopher” (34.38% missed), and “Mule” (28.91% missed).

### Distracter false alarm performance

The proportion of false alarms committed compared to all possible false alarm stimuli are reported in Fig. 5. There was a significant interaction between period on watch and false alarm group, $F(8, 242) = 17.58$, $p < .001$, $\eta_p^2 = 0.22$, $\varepsilon = 0.75$. There was a significant main effect of period on watch, $F(4, 120) = 63.77$, $p < .001$, $\eta_p^2 = 0.34$, $\varepsilon = 0.75$, and a significant main effect of false alarm type, $F(4, 120) = 73.13$, $p < .001$, $\eta_p^2 = 0.54$.

Observers across all conditions committed significantly fewer distracter false alarms as a function of time. Observers committing a combination of lure and distracter false alarms committed significantly more distracter false alarms than the lure only false alarm group ($p < .001$) and distracter only false alarm group (i.e., standard task; $p < .001$). Observers in the distracter false alarm only group committed the fewest distracter false alarms throughout the vigil.

### Distracter length effects

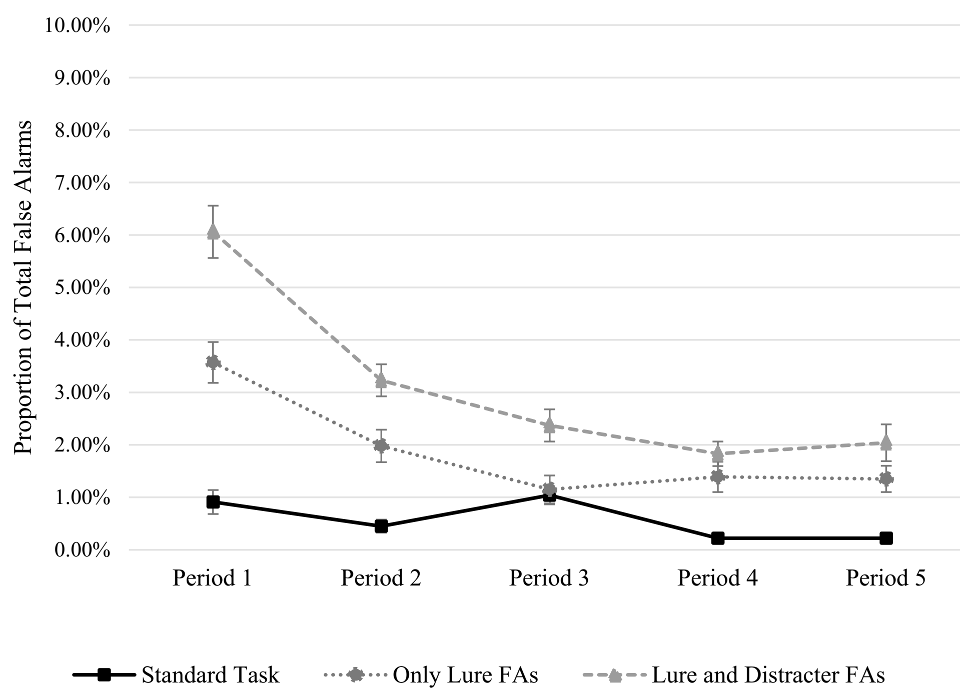
The average distracter word length is 4.9 letters in length (range 3–9 letters). Because so few distracters were committed (approximately 6% of all neutral events; see Fig. 5), we refrained from analyzing these results. However, there were several common distracter false alarms committed across groups. These included “Money” (9 total responses), “Beer” (5 total responses), “Jinx” (4 total responses), “Bomb” (3 total responses), and “Cane” (3 total responses). It is possible that due to the quick presentation of these stimuli (i.e., 200 ms) that some distracters were perceived as four-legged creatures. For example, money could easily be perceived as monkey, and beer as bear, or jinx as lynx. It is also possible that these distracters have high emotionality attached to them, and in a way, acted more closely to a lure. A complete list of all distracter responses recorded is included in the “Appendix”.

### Lure false alarm performance

The proportion of lure only false alarms out of all possible lure false alarms per period are reported in Fig. 6. There was a significant interaction between period on watch and lure false alarm group, $F(4, 228) = 2.63$, $p = .04$, $\eta_p^2 = 0.04$, $\varepsilon = 0.92$. There was a significant main effect of period on watch, $F(4, 54) = 51.67$, $p < .001$, $\eta_p^2 = 0.48$, $\varepsilon = 0.92$, and a significant main effect of false alarm type, $F(1, 57) = 4.53$, $p < .001$, $\eta_p^2 = 0.07$.

A significant decline in the proportion of lure false alarms was observed between Periods one to three, but an increase in the proportion of lure false alarms was demonstrated for all subsequent periods on watch for both false alarm groups. Pairwise comparisons indicated that the lure only false alarm group committed significantly fewer lure false alarms over time than observers who committed a combination of lure and distracter false alarms ($p = .04$).

**Fig. 5** Proportion of all possible distracter false alarms as a function of period on watch and false alarm group. Error bars represent the standard error of the mean



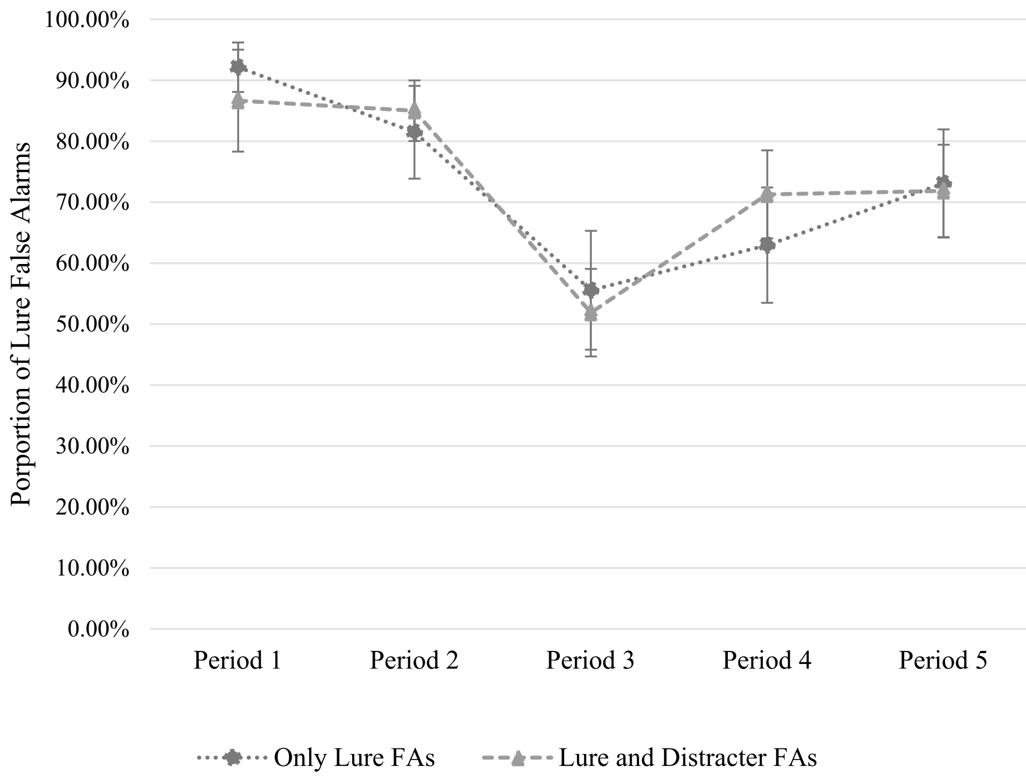



**Fig. 6** Proportion of lure only false alarms committed out of all possible lure false alarms per period across the lure false alarm groups. Error bars represent the standard error of the mean

### Lure length effects

The effects of lure word length are included in Fig. 7. The average lure word length is 5.44 letters in length (range 3–9 letters). Separate *t*-tests with Bonferroni corrections were performed for all possible lure letter lengths (i.e., 3–9 letters total in length). There was a significant main effect of false alarm group for three letter lure words, $t(57) = -2.70$, $p = .009$. Observers in the combination of lure and distracter false alarm group committed significantly more responses to three letter lure words than observers in the lure only false alarm group. There were no additional significant results to report for these analyses.

However, there were several common lure false alarms committed across groups. These included "Ant" (27 total responses), "Centipede" (22 total responses), "Cobra" (20 total responses), "Crab" (26 total responses), "Dolphin" (25 total responses), "Duck" (41 total responses), "Emu" (31 total responses), "Flamingo" (33 total responses), "Goose" (24 total responses), "Penguin" (19 total responses), "Rooster" (27 total responses), and "Turkey" (31 total responses). Again, it is possible that the short duration of presentation results in the misperception or misrepresentation of a word in memory.

### Response time

Average response time is reported in Fig. 8. There was a significant main effect of period on watch, $F(4, 120) = 32.97$, $p < .001$, $\eta_p^2 = 0.21$, $\varepsilon = 0.73$. Response time increased as a function of time on task across all five periods of watch.

There was also a significant main effect of false alarm type, $F(2, 123) = 33.90$, $p < .001$, $\eta_p^2 = 0.36$, which indicated that observers in the standard task (i.e., distracter only false alarms) responded more quickly than observers assigned to the lure condition, regardless of false alarm type. Response time was similar when lure only false alarms or a combination of lure and distracter false alarms were committed.

## Signal detection analyses

Mixed-measures ANOVAs were performed for the following measures of signal detection theory using period on watch as the within-subjects factor and false alarm type (i.e., standard task [distracter false alarms are the only false alarms possible], lure only false alarms, lure and distracter false alarms) as the between-subjects factor to identify significant differences over time for sensitivity and response bias.

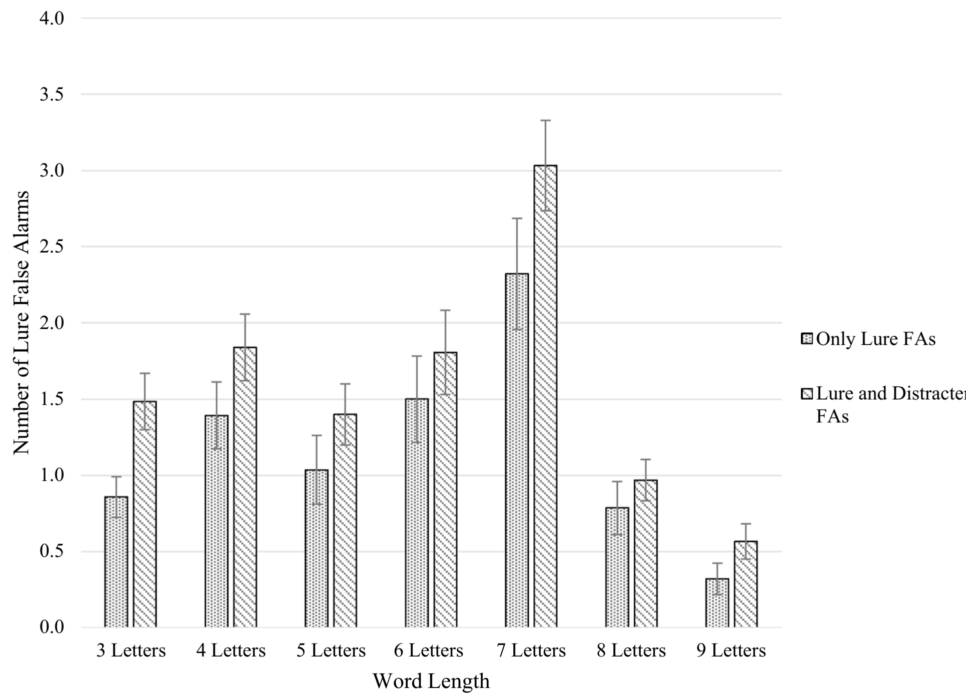



**Fig. 7** Average number of lure false alarms committed by false alarm group. Error bars represent the standard error of the mean. Note there are 4 three letter lure words, 5 four letter lure words, 12 five letter lure words, 10 six letter lure words, 10 seven letter lure words, 7 eight letter lure words, and 2 nine letter lure words included in the lure stimulus list

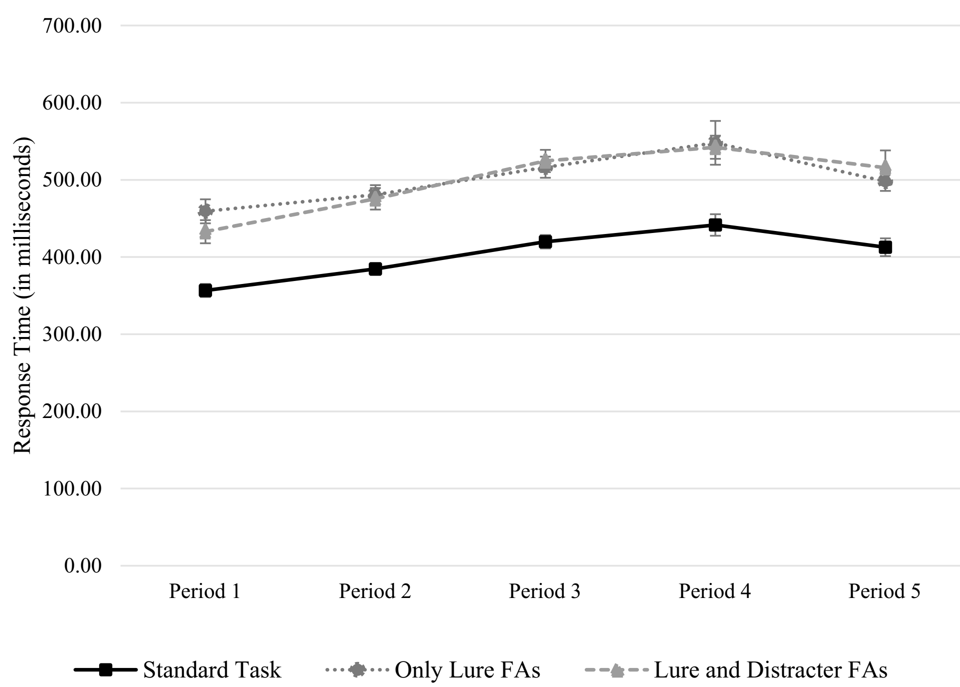



**Fig. 8** Response time as a function of period on watch and across false alarm groups. Error bars represent the standard error of the mean

Sensitivity was calculated using the proportion of correct detections and total proportion of false alarms using parametric *d'* (Green and Swets 1966; Macmillan and Creelman 2005). Response bias was calculated using the proportion of correct detections and total proportion of false alarms using parametric *c* (Green and Swets 1966; Macmillan and Creelman 2005). In some instances, the assumptions of sphericity were violated and a Huynh-Feldt epsilon statistic is, therefore, reported when appropriate.

## Sensitivity

Sensitivity (*d'*) is reported in Fig. 9. There was a significant main effect of period on watch, $F(4, 120) = 8.28$, $p < .001$, $\eta_p^2 = 0.06$, $\varepsilon = 0.84$, which indicated that perceptual sensitivity tended to increase with time on task. There was also a significant main effect of false alarm type on sensitivity, $F(2, 123) = 37.53$, $p < .001$, $\eta_p^2 = 0.38$, which indicated that observers in the distracter only group were more sensitive to stimuli than observers in the lure false alarm groups.

The pairwise comparisons indicated that observers in the distracter only false alarm group were significantly more sensitive than observers who committed a combination of lure and distracter false alarms ($p < .001$), as well as observers who committed lure false alarms only ($p < .001$). The latter lure false alarms groups were also significantly different from one another ($p < .001$).

### Response bias

Response bias ($c$) is reported in Fig. 10. There was a significant interaction between period on watch and lure false alarm group, $F(8, 242) = 7.22$, $p < .01$, $\eta_p^2 = 0.11$, $\varepsilon = 0.80$. There was a significant main effect of period on watch, $F(4, 120) = 42.44$, $p < .001$, $\eta_p^2 = 0.26$, $\varepsilon = 0.80$, and a significant main effect of false alarm type, $F(2, 123) = 4.06$, $p = .02$, $\eta_p^2 = 0.06$.

A significant difference between the groups was observed between Periods one to four; but toward the end of Period five, the groups approached similar levels of conservatism. Pairwise comparisons indicated that the distracter only false alarm group demonstrated significantly different trends in response bias compared to the lure only false alarm group ($p = .04$), and observers who committed a combination of lure and distracter false alarms ($p = .01$). However, response bias was not significantly different between either of the lure false alarm groups.

## Discussion

The results indicate unique lexical processing shifts over the course of a vigil in the presence of lure stimuli, which in turn affects observer decision-making criteria, as well as perceived stress and workload. By examining vigilance performance based on the types of false alarms committed during the task, we demonstrated that this method of parsing the data yields insight into the effects of lures on signal detection indices, such as sensitivity and response bias.

Perceptual sensitivity was highest for observers in the standard task and lowest for observers who committed lure and distracter false alarms. Such trends may be due to observers' disagreement concerning what is considered a "leg." For example, some individuals may consider a large flipper on a walrus to be a leg, as it has leg-like bones in its lower mass (note: 25% of observers in the lure task responded to 'walrus' as a critical signal). The meaning of "leg" is somewhat unclear and thus affects the decision-making criteria observers utilize to respond to stimuli (i.e., do I believe this creature has four legs?), thereby influencing sensitivity indices of vigilance performance. It is also possible that the additional processing

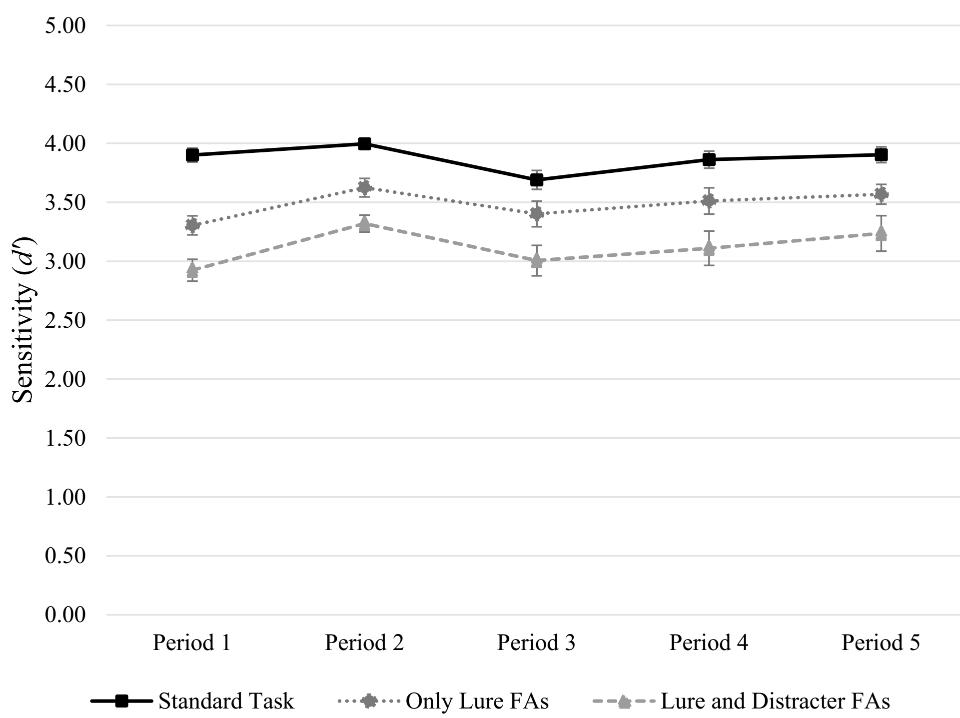



**Fig. 9** Sensitivity ($d'$) as a function of period on watch and false alarm group. Error bars represent the standard error of the mean

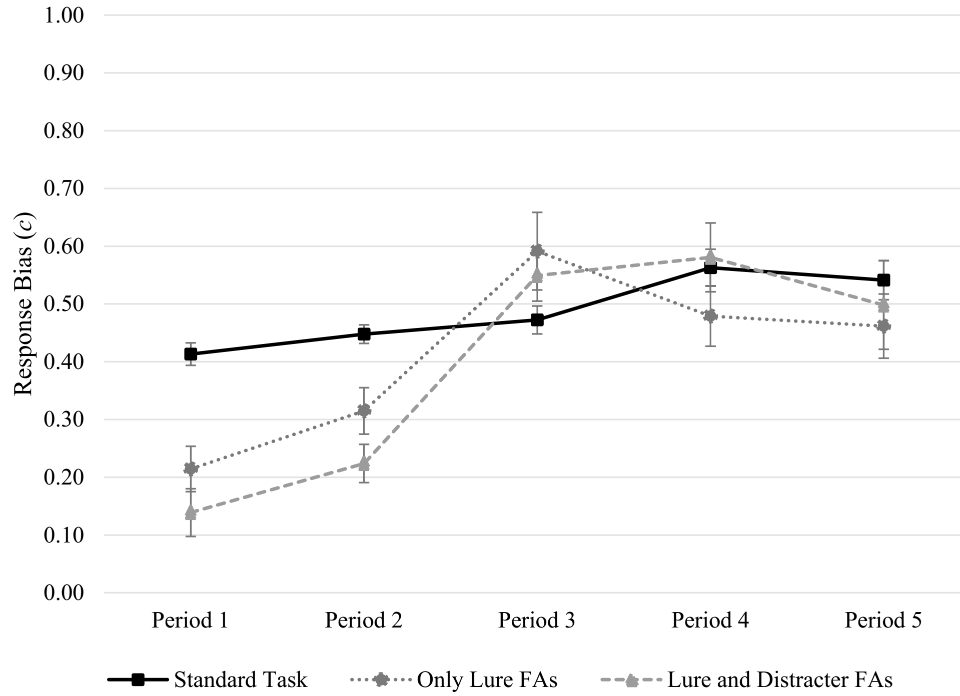



**Fig. 10** Response bias ($c$) as a function of period on watch and false alarm group. Error bars represent the standard error of the mean

requirements of the lure condition reduces the perceptual sensitivity toward the target stimuli.

Response bias indices were also significantly different between lure groups and observers in the standard task. While all groups indicated a shift toward conservatism as time on task increased, there were significant differences among the response criteria for the lure only false alarm group and the group that indicated a combination of lure and distracter false alarms. Both lure groups responded liberally until approximately the third period of the vigil.

Importantly, the analyses indicated that lexical processing of targets shifts in parallel with decision-making. Observers must access stored mental representations of each creature and ask themselves if they believe, or recall, the creature having four legs. The results demonstrated that observers committing lure false alarms likely use a multiple decision-making criterion strategy to decide if the stimulus is a signal (e.g., Is the stimulus a creature? Do I believe this creature is four-legged?). In contrast, observers committing both lure and distracter false alarms likely use a single criterion strategy (e.g., Is the target present?).

## Limitations and future directions

By analyzing the results in terms of the types of false alarm responses committed by each false alarm group, we discovered that performance could suffer due to the lexical meaning associated with target words. However, whether this is due to the increase in information processing afforded by the additional decision-making criteria required to respond to a target stimulus, to the affective processing instigated by the emotional meaning tied to any particular word, or a lack of common agreement over what is considered a leg, is yet to be determined.

In summation, the present study demonstrated that lexical vigilance tasks are significantly affected by the meaning attached to the definition of a target, which consequently affects indices of performance and signal detection. This research exemplifies the need to clearly define the concept of a critical signal for observers. On the other hand, the results also imply that the concept of a target in lexical vigilance tasks could be influenced by individual differences in the conceptualization of the target (i.e., categorically fuzzy or difficult to universally define). Future research on lexical vigilance tasks should seek to understand what exactly is driving disparities in performance in the presence of lure, or categorically similar, stimuli.

**Acknowledgements** We would like to thank our reviewers for providing valuable feedback on this manuscript, which served to make it more robust and interesting.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Disclosure statement** The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the

United States Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## Appendix

See Table 1.

**Table 1** Distracter false alarm responses

| Distracter stimulus | Number of responses |
|---|---|
| Alley | 2 |
| Art | 1 |
| Atom | 1 |
| Axe | 1 |
| Badge | 1 |
| Bagel | 1 |
| Beer | 5 |
| Belt | 1 |
| Bill | 2 |
| Binder | 1 |
| Bite | 1 |
| Blanket | 1 |
| Bomb | 3 |
| Bone | 2 |
| Bowl | 1 |
| Bracelet | 1 |
| Brick | 1 |
| Bulb | 1 |
| Bus | 1 |
| Cage | 1 |
| Candy | 1 |
| Cane | 3 |
| Canvas | 1 |
| Card | 1 |
| Cart | 1 |
| Cereal | 1 |
| Chair | 1 |
| Charger | 1 |
| Coast | 1 |
| Court | 1 |
| Crowd | 1 |
| Cube | 1 |
| Deed | 1 |
| Desk | 1 |
| Disc | 1 |
| Door | 1 |
| Drive | 1 |
| Element | 1 |
| Farm | 2 |
| Flag | 1 |
| Frame | 1 |
| Game | 2 |
| Jinx | 4 |
| Leap | 2 |
| Leash | 1 |
| Mine | 1 |
| Money | 9 |
| Movie | 1 |
| Nail | 1 |
| Needle | 1 |
| Net | 1 |
| Paint | 1 |
| Paste | 1 |
| Path | 1 |
| Plate | 1 |
| Puck | 1 |
| Pump | 1 |
| Radio | 1 |
| Sail | 1 |
| Screen | 1 |
| Slipper | 1 |
| Soap | 1 |
| Socket | 1 |
| Sole | 1 |
| Street | 1 |
| Sugar | 1 |
| Tank | 1 |
| Tent | 1 |
| Toy | 1 |
| Tray | 1 |
| Wave | 1 |
| Wire | 1 |
| Zipper | 1 |

## References


Baker CH (1959) Towards a theory of vigilance. Can J Exp Psychol 13:35–42

Claypoole VL, Neigel AR, Fraulini NW, Hancock GM, Szalma JL (2017) Can vigilance tasks be administered online? A replication and discussion. J Exp Psychol Hum Percept Perform

Craig A (1978) Is the vigilance decrement simply a response adjustment towards probability matching? Hum Factors 20:441–446

Craig A (1987) Signal detection theory and probability matching apply to vigilance. Hum Factors 29:645–652

Davies DR, Parasuraman R (1982) The psychology of vigilance. Academic Press, London

Deaton JE, Parasuraman R (1993) Sensory and cognitive vigilance effects of age on performance and subjective workload. Hum Perform 6:71–97

Deese J (1955) Some problems in the theory of vigilance. Psychol Rev 62:359–368

Epling SL, Russell PN, Helton WS (2016) A new semantic vigilance task: vigilance decrement, workload, and sensitivity to dual-task costs. Exp Brain Res 234:133–139

Frankmann JP, Adams J (1962) Theories of vigilance. Psychol Bull 59:257–272

Fraulini NW, Hancock GM, Neigel AR, Claypoole VL, Szalma JL (2017) A critical examination of the research and theoretical underpinnings discussed in Thomson, Besner, and Smilek (2016): a commentary. Psychol Rev 124:525–531

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley, New York

Grier RA, Warm JS, Dember WN, Matthews G, Galinsky TL, Szalma JL, Parasuraman R (2003) The vigilance decrement reflects limitations in effortful attention, not mindlessness. Hum Factors 45:349–359

Hart SG (2006) NASA-Task Load Index (NASA-TLX): 20 years later. Proc Hum Fact Ergo Soc Annu Meet 50:904–908

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol Res 52:139–183

Head J, Russel PN, Dorahy MJ, Neumann E, Helton WS (2012) Text-speak processing and the sustained attention to response task. Exp Brain Res 216:103–111

Head J, Wilson KM, Helton WS, Neumann E, Russell PN, Shears C (2013a) Right hemisphere prefrontal cortical involvement in text-speak processing. Proc Hum Fact Ergo Soc Annu Meet 57:379–383

Head J, Neumann E, Helton WS, Shears C (2013b) Novel word processing. Am J Psychol 126:323–333

Jerison HJ (1970) Vigilance, discrimination and attention. In: Mostofsky DI (ed) Attention: contemporary theory and analysis. Appleton, New York, pp 127–147

Mackworth NH (1948) The breakdown of vigilance during prolonged visual search. Q J Exp Psychol 1:6–21

Macmillan NA, Creelman CD (2005) Detection theory. London Erlbaum Associates, New Jersey

Majtásová L, Šípoš I (1974) Vigilance in acoustic-semantic signals with and without meaningful continuity. Stud Psychol 16:214–216

Matthews G (2016) Multidimensional profiling of task stress states for human factors: a brief review. Hum Factors 58:801–813

Matthews G, Campbell SE, Falconer S et al (2002) Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion 2:315–340

Parasuraman R, Mouloua M (1987) Interaction of signal discriminability and task type in vigilance decrement. Percept Psychophys 41:17–22

Shaw TH, Satterfield K, Ramirez R, Finomore V (2013) Using cerebral hemovelocity to measure workload during a spatialised auditory vigilance task in novice and experienced observers. Ergonomics 56:1251–1263

Thomson DR, Hasher L (2017) On the preservation of vigilant attention to semantic information in healthy aging. Exp Brain Res 235:2287–2300

Thomson DR, Besner D, Smilek D (2016) A critical examination of the evidence for sensitivity loss in modern vigilance tasks. Psychol Rev 123:70–83

Vickers D (1979) Decisions processes in visual perception. Academic Press, London

Warm JS (1977) Psychological processes in sustained attention. In: Mackie RR (ed) Vigilance: theory, operational performance and physiological correlates. Plenum, New York, pp 623–644

Xiuying Q, Junying Z (1998) The vigilance performance of Chinese signal monitoring. CNKI 1:1671

Yap M, Seow CS (2014) The influence of emotion on lexical processing: insights from RT distributional analysis. Psychon Bull Rev 21:526–533