False Feedback Increases Detection of Low Prevalence Targets in Visual Search

Jeremy Schwark, Joshua Sandry, Justin MacDonald, and Igor Dolgov

New Mexico State University

Word Count: 2415

Author Note

Jeremy Schwark, Joshua Sandry, Justin MacDonald, and Igor Dolgov, Department of Psychology, New Mexico State University, Las Cruces, NM.

Correspondence should be addressed to Jeremy Schwark; jschwark@nmsu.edu, Department of Psychology, MSC 3452, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003-8001. Phone: (575) 520-5484, Fax: (575) 646-6212.

The second author was supported in part by the National Science Foundation (NSF) through the Graduate STEM Fellows in K-12 Education (GK-12) Program, under grant number DGE-0947465. The authors would like to thank Robert Harris for his help creating the stimuli and Harris Schwark for his comments.

Abstract

Many critical search tasks, such as airport and medical screening, involve searching for targets that are rarely present. These low prevalence targets are associated with extremely high miss rates (Wolfe, et al., 2005). The inflated miss rates are caused by a criterion shift, likely due to observers attempting to equate the number of misses and false alarms. This equalizing strategy results in a neutral criterion at 50% target prevalence, but leads to a higher proportion of misses for low prevalence targets. The current study manipulated participants' perceived number of misses through explicit, false feedback. As predicted, participants in the false feedback condition committed a higher number of false alarms due to a shifted criterion. Importantly, participants in the false feedback condition were also more successful in detecting targets. These results highlight the importance of perceived prevalence in target search tasks.

Keywords: Signal Detection Theory; Visual Search; Low Prevalence; Feedback

False Feedback Increases Detection of Low Prevalence Targets in Visual Search

Visual search tasks are something we partake in every day. While some searches are rather trivial in nature (looking for the shirt we want to wear, finding food in the refrigerator, locating our car keys, etc.) other search tasks play a vital role in our wellbeing. Airport security, radiologists, and military personnel all perform critical, visual search tasks that can have serious repercussions if the targets that they are searching for are not detected.

These critical search tasks are more difficult when targets are rare (i.e., have a low prevalence rate), as is often the case. The likelihood of missing a target is substantially higher for low prevalence targets; a finding termed the low prevalence (LP) effect (Wolfe, Horowitz, & Kenner, 2005). Wolfe and colleagues found that target miss rates were only 7% when a target appeared in 50% of the trials, but increased to 30% when a target appeared in only 1% of the trials. This effect has serious implications in a critical search task such as medical screening, where the prevalence of a target can be less than 1% (Fenton et al., 2007).

Analyses using signal detection theory (Green & Swets, 1967) revealed that the LP effect was the result of a criterion shift rather than a loss in sensitivity (Wolfe et al., 2007; Wolfe & Van Wert, 2010). As the prevalence of a target decreases, observers become biased against the 'target detected' response. Some evidence also suggests that a speed-accuracy tradeoff could contribute to the LP effect (Fleck & Mitroff, 2007), but further research revealed that the speed-accuracy tradeoff was primarily responsible for misses due to motor-response errors, not misses resulting from a criterion shift (Rich et al, 2008; Van Wert, Horowitz, & Wolfe, 2009).

Wolfe and colleagues (2007) modeled the LP effect based on three assumptions. The first was that changes in target prevalence shift criterion, not sensitivity, as demonstrated by signal detection theory. The second was that observers attempt to equate the number (rather than the proportions) of false alarms and misses that they commit. This strategy results in a neutral criterion at 50% prevalence, but at low prevalence it leads to a high proportion of misses due to the rarity of the target event. The third assumption was that observers operate on an ROC function with a slope less than 1. The current study focuses on the second assumption and the role of feedback on errors in low prevalence search.

Observers use feedback in order to set their decision criterion in visual search tasks (Chun & Wolfe, 1996). Feedback can either be implicit (the observer *knows* they were correct because they saw the target) or explicit (the observer is told that they were correct or incorrect). Explicit feedback has been shown to have an impact on the LP effect. For example, providing feedback during bursts of high prevalence trials and no feedback during low prevalence trials allows observers to form and maintain a good criterion, counteracting the LP effect (Wolfe et al., 2007). Van Wert, Horowitz, and Wolfe (2009) found the LP effect was present, but slightly diminished, when observers were not given any feedback.

If observers use feedback to adjust their criterion in order to equate misses and false alarms, it may be possible to manipulate their criterion by adjusting the explicit feedback that they receive. When observers are informed that they are committing more misses than they actually are, they should shift their criterion in a way that would cause them to commit more false alarms and, more importantly, cause them to correctly identify more targets.

Experiment 1

Experiment 1 sought to test whether misleading, explicit feedback could shift observers' criterion and cause them to identify more targets. In 20% of the trials the participants were falsely informed that they had missed a target even though they had correctly indicated "target absent". We hypothesized that, if participants equate the number of misses and false alarms that they commit, increasing the participants' perceived number of misses should lead to an increase in perceived number of targets and cause them to adjust their criterion so that they commit more false alarms and detect more targets.

Method

Twenty undergraduate students (10 female, 10 male) from New Mexico State University participated for partial course credit. The mean age was 19.5 years (*SD* = 2.7). All participants had normal or corrected-to-normal vision.

The experiment was run using E-Prime 2.0 on a computer with a 21" monitor set at 1920 x 1080 resolution and a refresh rate of 65 Hz at a distance of approximately 20". Stimuli were created by randomly placing an array of 300 letters in a 400 x 400 pixel area (Figure 1). Capital letters were presented in black 16 pt. Arial font on a white background. In target-present stimuli, a single letter X appeared in the image and in target-absent stimuli there was no letter X. Letters were permitted to overlap. These stimuli were designed to be challenging so that the false feedback would be more believable.

Participants were informed that they would be completing a target search task in which they would be asked to indicate whether the letter X was present in an image. Additionally, participants were told that they would be accumulating points for correct answers and losing points for incorrect answers and their goal was to accumulate the most points. Trials started with a fixation cross (500 ms), followed by the stimulus. The stimulus was displayed until an answer was provided or until 10s had elapsed. Participants pressed "j" to indicate target present and "f" to indicate target absent, followed by the feedback screen (1500 ms) indicating whether the participant's choice was correct or incorrect.

The scoring system used was identical to that used by Wolfe et al. (2007), in which participants lost 75 points for false alarms and 150 points for misses and gained 25 points for hits and 5 points for correct rejections. However, due to the difficulty of the task, participants started with 10,000 points to reduce the likelihood that they would see a negative point total on the feedback screen. The participant’s running point total was displayed on the feedback screen.

Participants were randomly assigned to a feedback condition that determined whether they received true or false feedback in the low prevalence block. In the true feedback condition, participants always received accurate feedback. In the false feedback condition, participants received false feedback on 20% of the trials (target-absent trials only). False feedback was given only if the participant correctly reported “target absent” on the trial. False feedback looked identical to the actual “miss” feedback.

All participants completed 20 practice trials at 50% prevalence, 50 experimental trials at high prevalence (50% prevalence), and 200 experimental trials at low prevalence (4% prevalence). The high and low prevalence blocks were counterbalanced across participants. False feedback only occurred during the low prevalence trials. Participants were given an optional break after every 50 trials. The experiment took approximately 1 hour to complete.

Results and Discussion

For all analyses, false feedback trials were not included in any of the miss rates (miss rates were calculated only from actual target-present trials). A paired samples t-test revealed a pronounced LP effect, with target miss rates significantly higher in the low prevalence block than the high prevalence block across conditions, $t(19) = 4.00$, $p < .001$, $d = 1.46$. An independent samples t-test performed between the true feedback and false feedback conditions in the low prevalence trials revealed a significant difference in miss rates, $t(18) = 3.91$, $p < .001$, *d*

= 1.84, and false alarm rates, $t(18) = 3.56$, $p < .01$, $d = 1.68$. Miss rates in the false feedback condition were significantly lower than in the true feedback condition and false alarm rates were higher in the false feedback condition than in the true feedback condition (see Figure 2 for all comparisons).

These results demonstrate that participants found more LP targets in the false feedback condition while also committing more false alarms. A signal detection theory analysis[1] revealed that false feedback also significantly shifted participants' criterion (*C*), $t(13) = 4.01$, $p < .001$, $d = 2.36$, but did not significantly change sensitivity (*d'*), $p = .22$ (see Table 1). These results are consistent with previous findings that changes in LP event detection are due to shifts in criterion rather than sensitivity (Wolfe et al., 2007).

These data support the hypothesis that participants detect more targets if their perceived miss rate is inflated through false feedback. However, the overall miss and false alarm rates in both the high and low prevalence blocks were quite high due to the difficulty of the task. We were initially concerned that, if the task was too easy, participants would not be misled by the false feedback and would be aware of the deception. In an attempt to avoid this, we designed the stimuli to be densely cluttered so that the false feedback would be more believable. However, this caused some participants' performance to encounter a ceiling effect in miss and false alarm rates.

Experiment 2

Experiment 2 sought to address the potential ceiling effect in Experiment 1 and reduce the overall miss and false alarm rates through the use of less densely cluttered stimuli. The hypothesis was identical to Experiment 1 and more trials were added to ensure that participants

[1] Some participants had to be removed from the SDT analyses due to a hit rate of 0% in the low prevalence block (n = 5 & n = 5, Experiments 1 & 2, respectively). Removal of these participants did not change the pattern of results in the miss and false alarm data.

were able to fully shift their criterion based on prevalence, although we do not expect the number of trials will influence the LP effects found in Experiment 1 (see Wolfe et al., 2007).

Method

Forty undergraduate students (23 female, 17 male) from New Mexico State University received partial course credit for participation. The mean age was 19.0 years (*SD* = 1.3). All participants had normal or corrected-to-normal vision.

The same materials from Experiment 1 were used, except the stimuli in Experiment 2 contained 100 letters instead of 300 (Figure 3). The procedure was identical to Experiment 1, except participants completed a total of 100 high prevalence trials and 300 low prevalence trials (counter balanced). The experiment took approximately 1 hour.

Results and Discussion

The same analyses were performed on the data from Experiment 2, replicating the pattern of results obtained in Experiment 1. There was a significant LP effect present across conditions, $t(39) = 6.51$, $p < .001$, $d = .98$. In the low prevalence trials, miss rates were lower in the false feedback condition than in the true feedback condition, $t(38) = 2.17$, $p < .05$, $d = .71$, and false alarms were higher in the false feedback condition than in the true feedback condition, $t(38) = 4.74$, $p < .001$, $d = 1.54$ (see Figure 2 all comparisons). As predicted, signal detection theory analyses revealed a significant shift in *C* across feedback conditions, $t(33) = 4.18$, $p < .001$, $d = 1.47$. Interestingly, *d'* did significantly decrease in the false feedback condition, $t(33) = 2.14$, $p = .04$, $d = .75$, although this finding was neither predicted nor supported by previous studies (see Table 1).

General Discussion

The data presented in this study showed that increasing observers' perceived misses through false feedback caused them to detect more low prevalence targets. Signal detection theory analysis revealed that, consistent with previous LP studies (Wolfe et al., 2007; Wolfe & Van Wert, 2010), this was caused by a shift in criterion. Also, observers committed more false alarms when their perceived miss rate was elevated through false feedback, providing evidence in support of the assumption that observers attempt to equate their number of misses and false alarms (Wolfe et al., 2007).

A simple way to eliminate the LP effect would be to "increase the prevalence to 50% by adding artificial target-present trials" (Wolfe et al., 2007, p. 624). However, as Wolfe and colleagues note, simply adding target-present trials would effectively double the number of trials and the length of the task, making the solution impractical. However, our results provide support for the notion that it may not be necessary change the *actual* prevalence of the target, but simply change the *perceived* prevalence of the target. Through the manipulation of feedback, an observer can be convinced that the target has a higher prevalence than it actually does, negating the need to extend the task. Of course, this method comes with serious drawbacks as well, such as the ethical issue of telling an observer that they have been consistently missing targets that are not actually there. Perhaps more importantly, the practicality of implementing false feedback in real-world vigilance tasks depends on the payoff matrix associated with the different response outcomes. While a reduction in the miss rate is always desirable, the increased false alarm rate that is likely to accompany it may render the false feedback method impractical. Assuming that observers set their criterion to match the *number* of false alarms and misses, however, a small increase in the false alarm rate will lead to a relatively large decrease in the miss rate.

Of course, the consistent similarity of the numbers of misses and false alarms may be an indirect consequence of a more complex decision strategy rather than a deliberate effort by the observer to make the numbers match as suggested by Wolfe and colleagues (2007). Instead of setting the criterion based on some monotonic transformation of the likelihood ratio as suggested by SDT, the Wolfe et al. (2007) model suggests that users somehow keep a mental tally of errors that are then used to place the criterion. This model does an admirable job of explaining the consistent empirical finding (e.g., Thomas, 1975; Kornbrot, Donnelly, & Galanter, 1981) that observers fail to shift their decision criterion as far as detection theory says they should in low prevalence conditions. However, given that observers also exhibit similar numbers of misses and false alarms in the absence of feedback (Wolfe et al., 2007), for the theory to hold observers must provide their own response feedback internally. This implies that observers make responses that they know to be incorrect. This assertion deserves further empirical investigation. In any case, the results of the current study indicate that external feedback overrides any internal error tally, and that the LP effect can be reduced or eliminated with inaccurate external feedback.

References

Chun, M. M. & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30*, 39-78.

Fenton, J. J., Taplin, S. H., Carney, P. A., Abraham, L., Sickles, E. A., D'Orsi, C., Berns, E. A., Cutter, G., Hendrick, R. E., Barlow, W. E., & Elmore, J. G. (2007). Influence of computer-aided detection on performance of screening mammography. *New England Journal of Medicine, 356*(14), 1399–1409.

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18*, 943–947.

Green, D. M., & Swets, J. A. (1967). *Signal detection theory and psychophysics*. New York: Wiley.

Kronbrot, D. E., Donnelly, M., & Galanter, E. (1981). Estiamtes of utility function parameters from signal detection experiments. *Journal of Experimental Psychology: Human Perception & Performance, 7,* 441-458.

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision, 8*, 1–17.

Thomas, E. A. C. (1975). Criterion adjustment and probability matching. *Perception & Psychophysics, 18,* 158-162.

Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics, 71*, 541-553.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435*, 439-440.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136*, 623–638.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology, 20*, 121–124.

Table 1

Signal detection theory means for Experiment 1 and 2. Standard deviations are presented in parentheses.

	Experiment 1			Experiment 2		
Variable	50%	4% True	4% False	50%	4% True	4% False
C	.21 (.52)	1.36 (.37)	.70 (.27)	.32 (.44)	1.16 (.38)	.59 (.41)
d'	.33 (.53)	1.11 (.53)	.65 (.66)	1.71 (1.06)	2.04 (.81)	1.54 (.56)

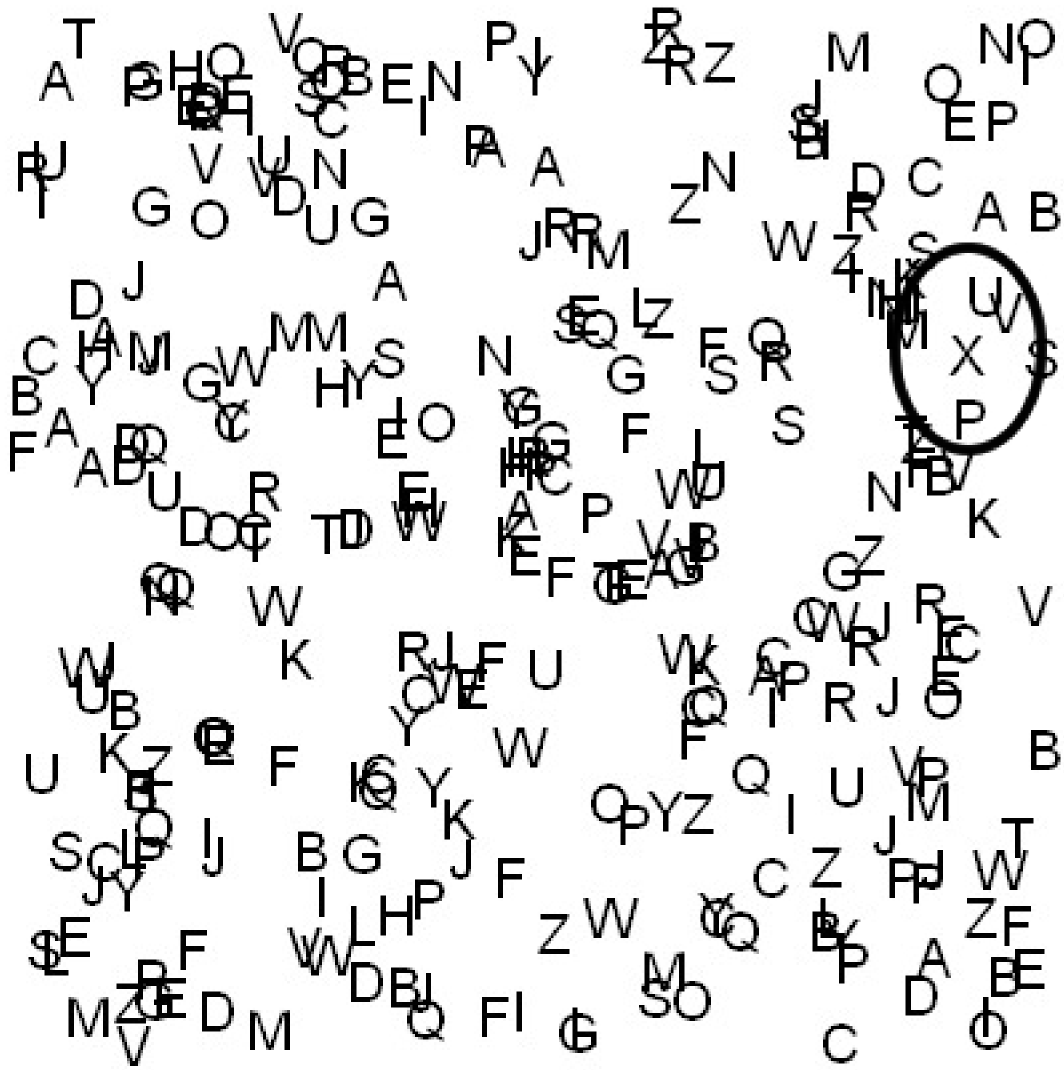

Figure 1. Sample target-present stimulus used in Experiment 1. The target letter (X) is circled for readability.

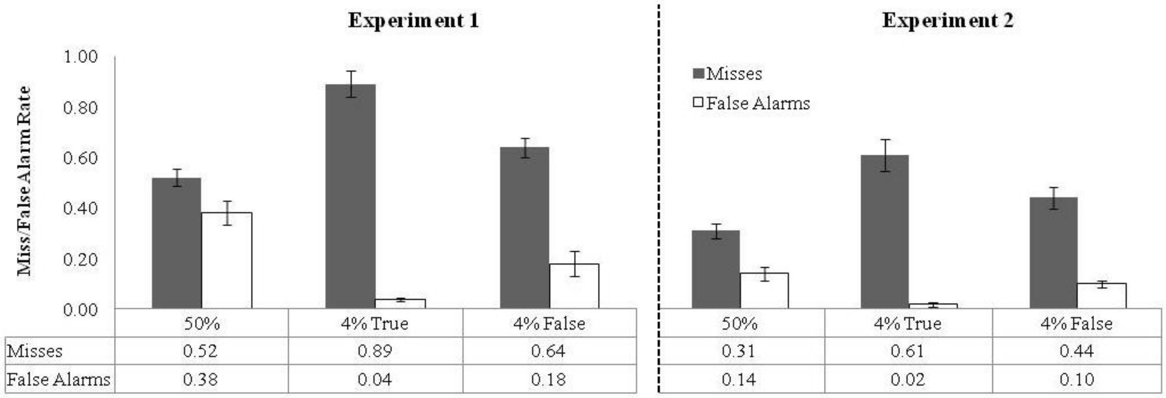

Figure 2. Miss and false alarm rates obtained in Experiments 1 & 2 show that participants missed significantly less targets in the false feedback condition (4% False) than the true feedback condition (4% True). Consequently, false alarm rates were higher in the false feedback condition than the true condition. Error bars are in standard error.

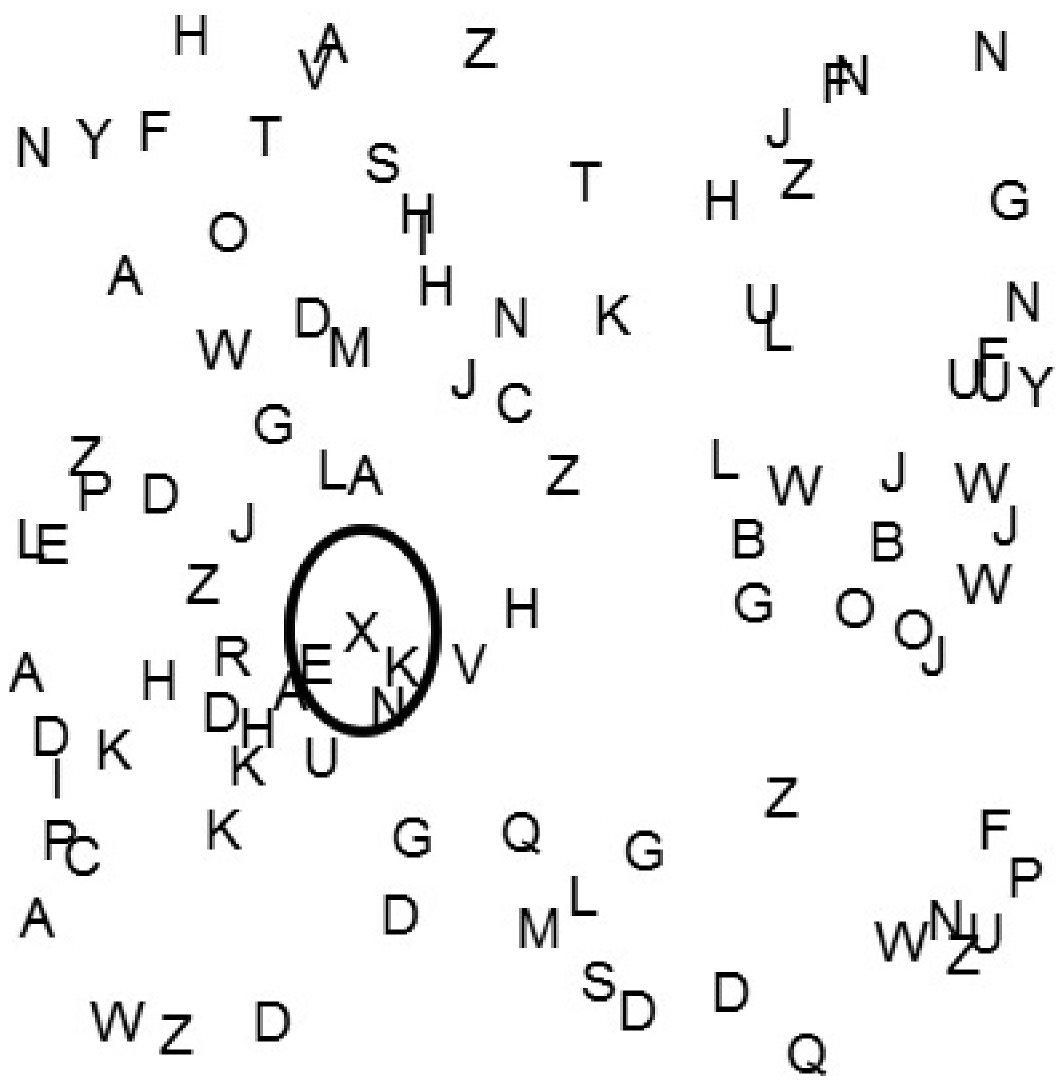

Figure 3. Sample target-present stimulus used in Experiment 2. The target letter (X) is circled for readability.