

# False feedback increases detection of low-prevalence targets in visual search

Jeremy Schwark · Joshua Sandry · Justin MacDonald · Igor Dolgov

Published online: 3 August 2012  
© Psychonomic Society, Inc. 2012

**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, Horowitz, & Kenner (Nature, 435, 439–440, 2005). The inflated miss rates are caused by a criterion shift, likely due to observers attempting to equate the numbers 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. In the present study, we manipulated participants' perceived number of misses through explicit false feedback. As predicted, the participants in the false-feedback condition committed a higher number of false alarms due to a shifted criterion. Importantly, the participants in this 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

Visual search tasks are something we partake in every day. While some searches are rather trivial in nature (e.g., looking for the shirt that we want to wear, finding food in the refrigerator, or locating our car keys), 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. Just recently, security officials at LAX failed to detect a loaded gun in a

handbag (Blankstein & Sewell, 2011) and caused a scare when they mistook an insulin pump for a gun (Blankstein, 2012). These critical search tasks are more difficult when the 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 et al. (2005) found that target miss rates were only 7 % when a target appeared in 50 % of the trials, but miss rates 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 (SDT; Green & Swets, 1966) have revealed that the LP effect is the result of a criterion shift rather than of 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 has also suggested that a speed–accuracy trade-off could contribute to the LP effect (Fleck & Mitroff, 2007), but further research revealed that the speed–accuracy trade-off 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 et al. (2007) modeled the LP effect on the basis of three assumptions. The first was that changes in target prevalence shift criterion, not sensitivity, consistent with SDT. The second was that observers attempt to equate the numbers (rather than the proportions) of false alarms and misses that they commit. This strategy results in an unbiased 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 in LP situations, observers operate on a receiver operating characteristic

J. Schwark (✉) · J. Sandry · J. MacDonald · I. Dolgov  
Department of Psychology, New Mexico State University,  
MSC 3452, P.O. Box 30001, Las Cruces, NM 88003-8001, USA  
e-mail: jschwark@nmsu.edu

(ROC) function with a slope less than 1. The present study focuses primarily on the second assumption and on the role that feedback plays for errors in low-prevalence search.

Observers use feedback in order to assess and adjust their performance in visual search tasks (Chun & Wolfe, 1996). Prior research has revealed that feedback (knowledge of the results), and even false feedback (false knowledge of results), serves to improve performance as compared to no feedback (Mackworth, 1964). Indeed, it is hard to imagine how an observer would incorporate base rate information to develop a response bias without making use of some type of feedback to determine how to shift the criterion. This feedback may increase sensitivity and reduce decision time (Micalizzi & Goldberg, 1989; but see also Szalma et al., 2006). Feedback can either be implicit (i.e., observers *know* that they were correct because they saw the target) or explicit (observers are 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 that the LP effect was present, but slightly diminished, when observers were not given any feedback as compared to full 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, to correctly identify more targets.

## Experiment 1

In Experiment 1, we 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 numbers of misses and false alarms that they commit, increasing the participants' perceived number of misses should lead to an increase in the perceived number of targets and cause them to adjust their criterion so that they commit more false alarms and detect more targets.

### Method

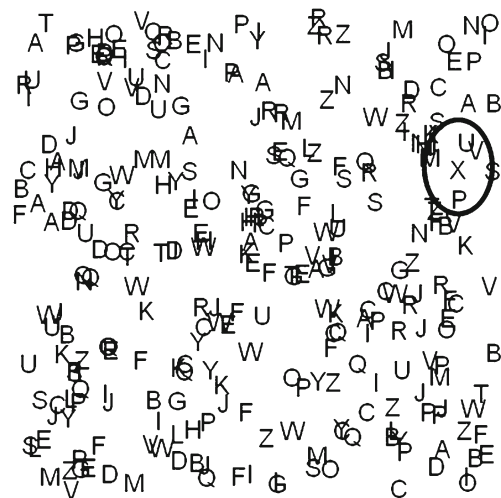
A group of 20 undergraduate students (10 female, 10 male) from New Mexico State University participated for partial

course credit. Their mean age was 19.5 years ( $SD = 2.7$ ). All participants had normal or corrected-to-normal vision.

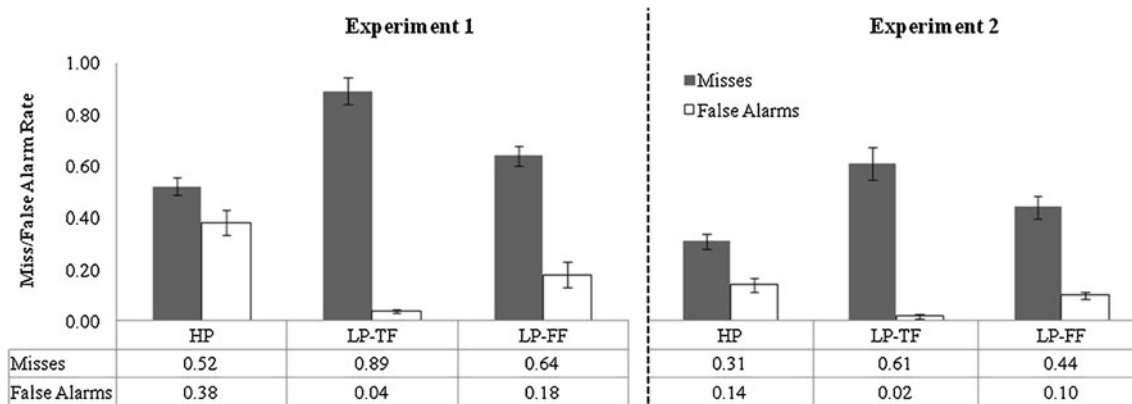
The experiment was run using E-Prime 2.0 on a computer with a 21-in. monitor set at  $1,920 \times 1,080$  resolution and a refresh rate of 65 Hz, at a distance of approximately 20 in. The stimuli were created by randomly placing an array of 300 letters in a  $400 \times 400$  pixel area (Fig. 1). Capital 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, the letter X was absent. The 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, the participants were told that they would be accumulating points for correct answers and losing points for incorrect answers, and that 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 10 s had elapsed. Participants pressed the "j" key to indicate "target present" and the "f" key to indicate "target absent," followed by a feedback screen (1,500 ms) indicating whether the participant's choice was correct or incorrect.

A scoring system was used to replicate applied situations in which hits are critically important and misses are extremely damaging. 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. Additionally, this scoring system worked *against* the LP effect by maximally discouraging misses. Due to the difficulty of the task and the false feedback, participants started



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



**Fig. 2** Miss and false alarm rates obtained in Experiments 1 and 2 show that participants missed significantly fewer targets in the false-feedback condition (LP-FF) than in the true-feedback condition (LP-TF).

Consequently, false alarm rates were higher in the false-feedback condition than in the true-feedback condition. Error bars represent standard errors

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, they 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 that trial. The 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 order of the high- and low-prevalence blocks was 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 h to complete.

**Results and discussion**

For all analyses, false-feedback trials were not included in the calculation of 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 in the high-prevalence block across conditions,  $t(19) = 4.00, p < 0.001$ , Cohen’s  $d = 1.46$ . An independent-samples *t* test performed between the true-feedback and false-feedback conditions in the low-prevalence trials revealed significant differences in miss rates,  $t(18) = 3.91, p < 0.001$ , Cohen’s  $d = 1.84$ , and false alarm rates,  $t(18) = 3.56, p < 0.01$ , Cohen’s  $d = 1.68$ . Miss rates in the false-feedback condition were significantly lower than those in the true-feedback condition, and

false alarm rates were higher in the false-feedback than in the true-feedback condition (see Fig. 2 for all comparisons).

These results demonstrate that participants found more LP targets in the false-feedback condition, while also committing more false alarms. An SDT analysis<sup>1</sup> revealed that false feedback also significantly shifted participants’ criterion (*C*; Macmillan & Creelman, 1990),  $t(13) = 4.01, p < 0.001$ , Cohen’s  $d = 2.36$ , toward a more neutral setting. However, false feedback did not significantly change sensitivity ( $d'$ ; Macmillan & Creelman, 1990),  $p = 0.22$  (see Table 1). These results are consistent with previous findings that have suggested that changes in LP event detection are due to shifts in criterion rather than sensitivity (Wolfe et al., 2007).

Confidence intervals (CIs) were used to determine whether participants were matching the numbers of misses and false alarms that they committed. In the high-prevalence block, overlapping confidence intervals found in the numbers of misses and false alarms, 95 % CIs [11.21, 14.69] and [6.89, 11.91] (miss CIs are presented before all false alarm CIs), show that participants were committing approximately the same number of each type of error. In the low-prevalence block, the numbers of misses and false alarms committed in the true-feedback condition, 95 % CIs [6.18, 8.02] and [5.39, 11.61], were also similar. In the false-feedback condition, it initially appeared that participants were committing fewer misses than false alarms, 95 % CIs [4.39, 5.81] and [18.14, 49.86]. However, when 40 misses were added to account for the false feedback (i.e., participants believed that they had committed misses on 40 trials), these confidence intervals also overlapped, 95 % CIs [44.39, 45.81] and [18.14, 49.86]. These results support the assumption made by Wolfe et al. (2007) that observers attempt to

<sup>1</sup> Five participants had to be removed from the SDT analyses due to hit rates of 0 % in the low-prevalence condition. Removal of these participants did not change the pattern of results in the miss and false alarm data.

**Table 1** Signal detection theory means for Experiments 1 and 2

Variable	Experiment 1			Experiment 2		
	50 %	4 % True	4 % False	50 %	4 % True	4 % False
<i>C</i>	0.21 (0.52)	1.36 (0.37)	0.70 (0.27)	0.32 (0.44)	1.16 (0.38)	0.59 (0.41)
<i>d'</i>	0.33 (0.53)	1.11 (0.53)	0.65 (0.66)	1.71 (1.06)	2.04 (0.81)	1.54 (0.56)

Standard deviations are in parentheses

match their misses and false alarms, as well as confirming our hypothesis that using false feedback to increase the perceived number of misses would increase the number of false alarms due to observers' attempts to match.

Analyses performed on RTs showed a speeding of “target absent” responses in the low-prevalence trials ( $M = 3,629$  ms,  $SD = 2,210$  ms) as compared to the high-prevalence trials ( $M = 4,562$  ms,  $SD = 2,234$  ms),  $t(19) = 2.48$ ,  $p = 0.023$ , Cohen's  $d = 0.43$ . These results are typical of the LP effect. However, “target absent” responses were not significantly slower in the false-feedback than in the true-feedback condition,  $t(18) = 1.60$ ,  $p = 0.13$ .

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 a floor effect in the performance of some of our participants.

## Experiment 2

Experiment 2 sought to address the potential floor effect in Experiment 1 and to reduce the overall miss and false alarm rates through the use of less densely cluttered stimuli. The hypothesis was identical to that of Experiment 1, and more trials were added to ensure that participants were able to fully shift their criterion on the basis of prevalence, although we did not expect the number of trials to influence the magnitude of LP effects (see Wolfe et al., 2007).

## Method

A group of 40 undergraduate students (23 female, 17 male) from New Mexico State University received partial course credit for participation. Their mean age was 19.0 years ( $SD = 1.3$ ). All participants had normal or corrected-to-normal vision.

The same materials were used as in Experiment 1, except that the stimuli in Experiment 2 contained 100 letters instead of 300 (Fig. 3). The procedure was identical to that of Experiment 1, except that participants completed a total of 100 high-prevalence and 300 low-prevalence trials (counterbalanced). The experiment took approximately 1 h to complete.

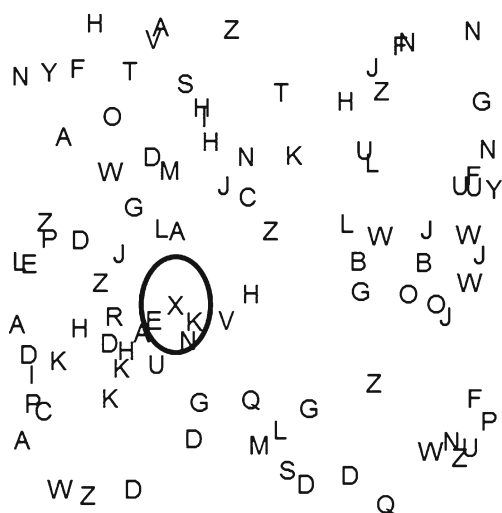
## Results and discussion

The same analyses were performed as on the data from Experiment 1, replicating the pattern of results obtained in the earlier experiment. A significant LP effect was present across conditions,  $t(39) = 6.51$ ,  $p < 0.001$ , Cohen's  $d = 0.98$ . In the low-prevalence trials, miss rates were lower in the false-feedback condition than in the true-feedback condition,  $t(38) = 2.17$ ,  $p = 0.037$ , Cohen's  $d = 0.71$ , and false alarm rates were higher in the false-feedback than in the true-feedback condition,  $t(38) = 4.74$ ,  $p < 0.001$ , Cohen's  $d = 1.54$  (see Fig. 2 for all comparisons).

As predicted, SDT analyses<sup>2</sup> revealed a significant shift in  $C$  across feedback conditions,  $t(33) = 4.18$ ,  $p < 0.001$ , Cohen's  $d = 1.47$ . Interestingly,  $d'$  did significantly decrease in the false-feedback condition,  $t(33) = 2.14$ ,  $p = 0.04$ , Cohen's  $d = 0.75$ , although this finding was neither predicted nor supported by previous studies (see Table 1).

Unlike in Experiment 1, the numbers of misses and false alarms that participants committed in the high-prevalence block were not quite equal, 95 % CIs [12.76, 18.24] and [4.12, 10.28]. Misses and false alarms were equal in the low-prevalence block with true feedback, 95 % CIs [5.71, 8.87] and [1.20, 12.62], but not in the block with false feedback (perceived misses added), 95 % CIs [64.09, 66.41] and [20.99, 36.61]. The reduction of false alarms as compared to misses likely occurred due to the reduced number of distractor letters used in the present experiment's stimuli, resulting in fewer occurrences of overlapping letters potentially appearing to be Xs. Still, the increased number of

<sup>2</sup> Five participants had to be removed from the SDT analyses due to hit rates of 0 % in the low-prevalence condition. Removal of these participants did not change the pattern of results in the miss and false alarm data.



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

false alarms in the false-feedback as compared to the true-feedback condition can be taken as evidence that the increase in the perceived number of misses had an impact on false alarms.

As in Experiment 1, analyses of RT data revealed the predicted speeding of “target absent” responses in the low-prevalence trials ( $M = 4,119$  ms,  $SD = 1,520$  ms) as compared to the high-prevalence trials ( $M = 5,057$  ms,  $SD = 2,023$  ms),  $t(39) = 3.05$ ,  $p = 0.004$ , Cohen’s  $d = 0.53$ . “Target absent” responses were not significantly slower in the false-feedback than in the true-feedback condition,  $t(38) = 0.05$ ,  $p = 0.96$ .

### Experiment 3

Experiment 3 was conducted to investigate the drops in  $d'$  found in Experiments 1 and 2 (significantly so, in Exp. 2). These findings are not typical of the LP effect, which is attributed to a change in criterion only (Wolfe et al., 2007), but Experiments 1 and 2 did not contain enough target-present trials to accurately plot ROC curves to fully investigate  $d'$ . Of interest was whether false feedback actually changed observers’  $d'$ .

To answer this question, in Experiment 3 we used four perceived target prevalence rates obtained through true feedback and false feedback. True-feedback blocks had actual (and assumed perceived) target prevalence rates of 24 %, 32 %, 40 %, and 48 %. False-feedback blocks had actual target prevalence rates of 4 %, 12 %, 20 %, and 28 %, with false feedback set at a constant 20 % in order to match the perceived prevalence rates of the true-feedback blocks when the false feedback was added to the real prevalence (e.g., 4 % real prevalence + 20 % false feedback = 24 % perceived

prevalence). Given that the stimulus set was held constant across all conditions, this procedure resulted in eight points with which to plot ROC curves in order to better estimate  $d'$  in the true- and false-feedback conditions. If  $d'$  was the same across true and false feedback, the plots from each feedback type should form a single curve. Experiment 3 also allowed us to test whether the zROC curves from the false-feedback blocks had slopes less than 1.0, which is also typical of the LP effect (Wolfe et al., 2007).

### Method

A group of 81 undergraduate students (47 female, 34 male) from New Mexico State University received partial course credit for participation. Their mean age was 19.5 years ( $SD = 2.1$ ). All participants had normal or corrected-to-normal vision.

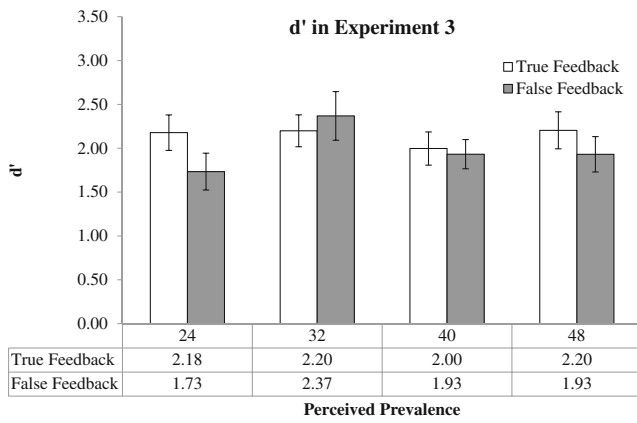
The same materials and stimuli were used as in Experiment 2. In order to sample additional prevalence points, extra stimuli were generated using the same parameters as in Experiment 2. Perceived prevalence was manipulated between participants at four levels: 24 % ( $n = 22$ ), 32 % ( $n = 20$ ), 40 % ( $n = 18$ ), or 48 % ( $n = 21$ ), and feedback veracity was blocked within participants. In the false-feedback block, the actual target prevalence was 4 %, 12 %, 20 %, or 28 % and participants received false feedback on 20 % of the trials (target-absent trials only), resulting in 24 %, 32 %, 40 %, or 48 % perceived prevalence, depending on the prevalence condition. All participants completed 20 practice trials at 50 % prevalence and then received 200 trials each of true and false feedback (blocked and counterbalanced) at their perceived target prevalence. The experiment took approximately 1 h to complete.

### Results and discussion

A  $2 \times 4$  mixed ANOVA<sup>3</sup> (Feedback Type  $\times$  Perceived Prevalence) using  $d'$  as the dependent variable revealed a significant Feedback Type  $\times$  Perceived Prevalence interaction,  $F(3, 70) = 2.83$ ,  $p = 0.044$ ,  $\eta_p^2 = 0.11$ . A follow-up test of the simple effects revealed a significant effect of feedback type in the 24 %-perceived-prevalence condition only,  $F(1, 70) = 7.94$ ,  $p = 0.006$ . All other simple effects of feedback type did not reach significance. As expected, there was no significant main effect of perceived prevalence on  $d'$  (Fig. 4),  $F(3, 70) = 0.62$ ,  $p = 0.607$ ,  $\eta_p^2 = 0.026$ .

In order to better understand the results, the data were graphed in ROC space by plotting hit and false alarm rates (Fig. 5). zROCs were also plotted, which revealed a slope of

<sup>3</sup> Seven participants were excluded from the analysis due to nonpositive  $d'$  values.

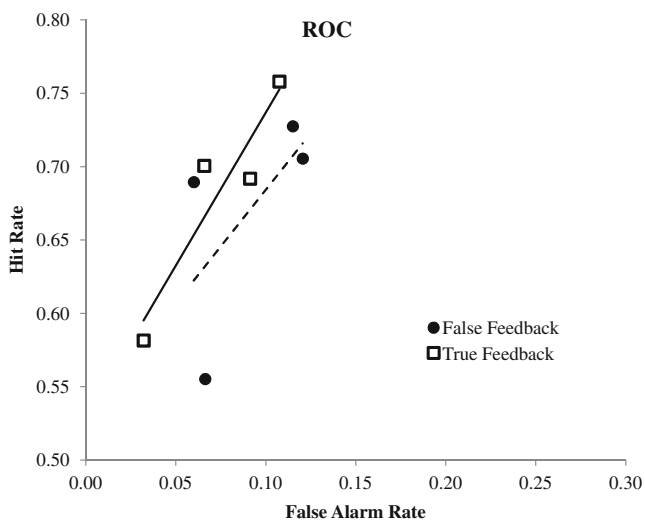


**Fig. 4** The  $d'$  measurements in Experiment 3 revealed no systematic difference between true and false feedback across the perceived-prevalence conditions. However, it is interesting to note that a significant difference was found in the lowest-prevalence condition. The error bars represent standard errors

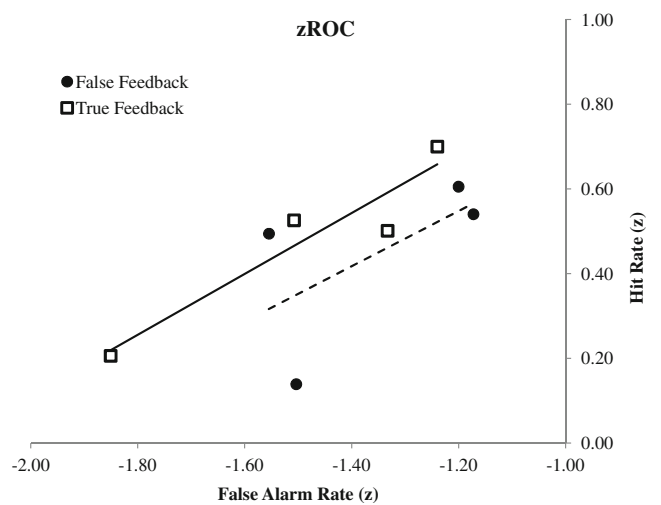
less than 1 for both the false-feedback (slope = 0.71) and true-feedback (slope = 0.65) conditions, a result typically found in LP studies (Fig. 6; Wolfe & Van Wert, 2010).

**General discussion**

The data from the present experiments showed that increasing observers' perceived misses through false feedback caused them to detect more low-prevalence targets. Consistent with previous LP studies (Wolfe et al., 2007; Wolfe & Van Wert, 2010), SDT measures indicated that this effect was caused by a shift in criterion. The criterion was closest to neutral in the 50 % block, was most conservative in the 4 % true-feedback condition, and less conservative in the 4 % false-feedback condition. The finding that observers committed more false



**Fig. 5** ROC graph plotted from Experiment 3. The dashed line is the fit line for false feedback



**Fig. 6** zROC graph plotted from Experiment 3. The dashed line is the fit line for false feedback

alarms when their perceived miss rate was elevated through false feedback provides strong evidence in support of the assumption that observers attempt to equate the numbers of misses and false alarms that they commit (Wolfe et al., 2007), although, as Experiment 2 showed, this attempt may only result in an approximate match between misses and false alarms. Even though some differences in sensitivity were found in Experiments 1 and 2 due to false feedback, Experiment 3 supports the notion that false feedback changes criterion rather than sensitivity. However, the significant difference in sensitivity found in the lowest-prevalence condition of Experiment 3 does open the door for the possibility that false feedback might have an impact on sensitivity at low target prevalence. However, further research will be necessary in order to determine whether this is a systematic effect or some failure of SDT to deal with false feedback.

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 et al. (2007) noted, 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 to change the *actual* prevalence of the target, but simply to 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. Although this was not tested in our study, false feedback should be able to improve performance at any target prevalence by shifting the criterion, even at a prevalence of 50 % or higher. However, an insufficiently set criterion is most likely to be problematic at lower target prevalence (Wolfe et al., 2007).

This method comes with serious drawbacks as well, such as the ethical issue of telling observers that they have

consistently missed targets that were 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 *numbers* 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 miss and false alarm rates 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 was suggested by Wolfe et al. (2007). Instead of setting the criterion on the basis of 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 is then used to place the criterion. This model does an admirable job of explaining the consistent empirical finding (e.g., Kornbrot, Donnelly, & Galanter, 1981; Thomas, 1975) that observers fail to shift their decision criterion as far as detection theory says that 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 present study indicate that external feedback overrides any internal error tally and that the LP effect can be reduced or eliminated with inaccurate external feedback.

**Author note** The second author was supported in part by the National Science Foundation through the Graduate STEM Fellows in K–12 Education Program, under Grant No. DGE-0947465. The authors thank Robert Harris for his help creating the stimuli, and Hayward Godwin, Jeremy Wolfe, Harris Schwark, and an anonymous reviewer for their comments on an earlier draft of the manuscript.

## References

- Blankstein, A. (2012, Jan. 27). TSA mistakes insulin pump for gun, causing LAX security scare. *Los Angeles Times*. Retrieved from <http://latimesblogs.latimes.com/lanow/2012/01/insulin-pump-lax-airport-security.html>
- Blankstein, A., & Sewell, A. (2011, Oct. 23). Security officials at LAX fail to detect loaded gun in bag. *Los Angeles Times*. Retrieved from <http://articles.latimes.com/2011/oct/23/local/la-me-tsa-handgun-20111024>
- 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. doi:10.1006/cogp.1996.0002
- Fenton, J. J., Taplin, S. H., Carney, P. A., Abraham, L., Sickles, E. A., D’Orsi, C., & Elmore, J. G. (2007). Influence of computer-aided detection on performance of screening mammography. *New England Journal of Medicine*, *356*, 1399–1409. doi:10.1056/NEJMoa066099
- 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. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Kornbrot, D. E., Donnelly, M., & Galanter, E. (1981). Estimates of utility function parameters from signal detection experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 441–458. doi:10.1037/0096-1523.7.2.441
- Mackworth, J. F. (1964). The effect of true and false knowledge of results on the detectability of signals in a vigilance task. *Canadian Journal of Psychology*, *18*, 106–117.
- Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and “nonparametric” indexes. *Psychological Bulletin*, *107*, 401–413. doi:10.1037/0033-2909.107.3.401
- Micalizzi, J., & Goldberg, J. H. (1989). Knowledge of results in visual inspection decisions: Sensitivity or criterion effect? *International Journal of Industrial Ergonomics*, *4*, 225–235.
- 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*(15), 1–17. doi:10.1167/8.15.15
- Szalma, J. L., Hancock, P. A., Warm, J. S., Dember, W. N., & Parsons, K. S. (2006). Training for vigilance: Using predictive power to evaluate feedback effectiveness. *Human Factors*, *48*, 682–692. doi:10.1518/001872006779166343
- 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. doi:10.3758/APP.71.3.541
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*, 439–440. doi:10.1038/435439a
- 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. doi:10.1037/0096-3445.136.4.623
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*, 121–124. doi:10.1016/j.cub.2009.11.066