Attentional Priority for Temporary Goals:

Evidence with a confound-free dot-probe task

Dirk Wentura

Benedikt Emanuel Wirth

Saarland University, Saarbrücken, Germany

Dirk Wentura: https://orcid.org/ 0000-0001-9907-498X

Benedikt Emanuel Wirth: https://orcid.org/0000-0001-8039-7316

Author Note

Both authors contributed equally to this work. This research was supported by a grant from the German Research Foundation to Benedikt Wirth and Dirk Wentura (WI 5758/1-1; WE 2284/16-1). Benedikt Wirth is now at the department of Cognitive Assistants at the German Research Center for Artificial Intelligence (DFKI). The data and the program code for all experiments are available on the Open Science Framework (OSF; Wentura & Wirth, 2023, August 16). Correspondence concerning this article should be addressed either to Dirk Wentura, E-mail: wentura@mx.uni-saarland.de or Benedikt Wirth, benedikt.wirth@dfki.de

Abstract

Numerous studies using the dot-probe task showed that stimuli associated with current goals (induced by instructions) are attentionally prioritized despite not sharing features with the search targets. However, the exact nature of this prioritization remains unclear. Because these studies employed a target location task, results can be interpreted both in terms of spatial attention and response-related processes; the target discrimination task – in contrast – allows unambiguous interpretation of effects in terms of spatial attention. In order to disentangle the mechanisms underlying prioritization of goal-related stimuli, we conducted two experiments. In Experiment 1, we replicated the original results of previous studies using the location task. In Experiment 2, we found a corresponding effect in the discrimination task, which provides an unambiguous interpretation of stimuli associated with current goals being prioritized with respect to spatial attention. A cross-experiments analysis indicates that the effect found with the discrimination task was significantly smaller than the one found with the location task. The reduced effect size in the discrimination task suggests that in the original experiments, response-related processes were at play in addition to attentional processes. In addition, the task manipulation partly clarified a surprising result found with the location task in previous studies and in Experiment 1: Threat-related stimuli showed no cueing effects if contrasted with neutral stimuli. In Experiment 2 (discrimination task), however, we found a negatively signed effect, a finding that could indicate attentional avoidance of threatening stimuli.

Keywords: Attentional bias; dot-probe task; spatial attention; goals; threat; response priming

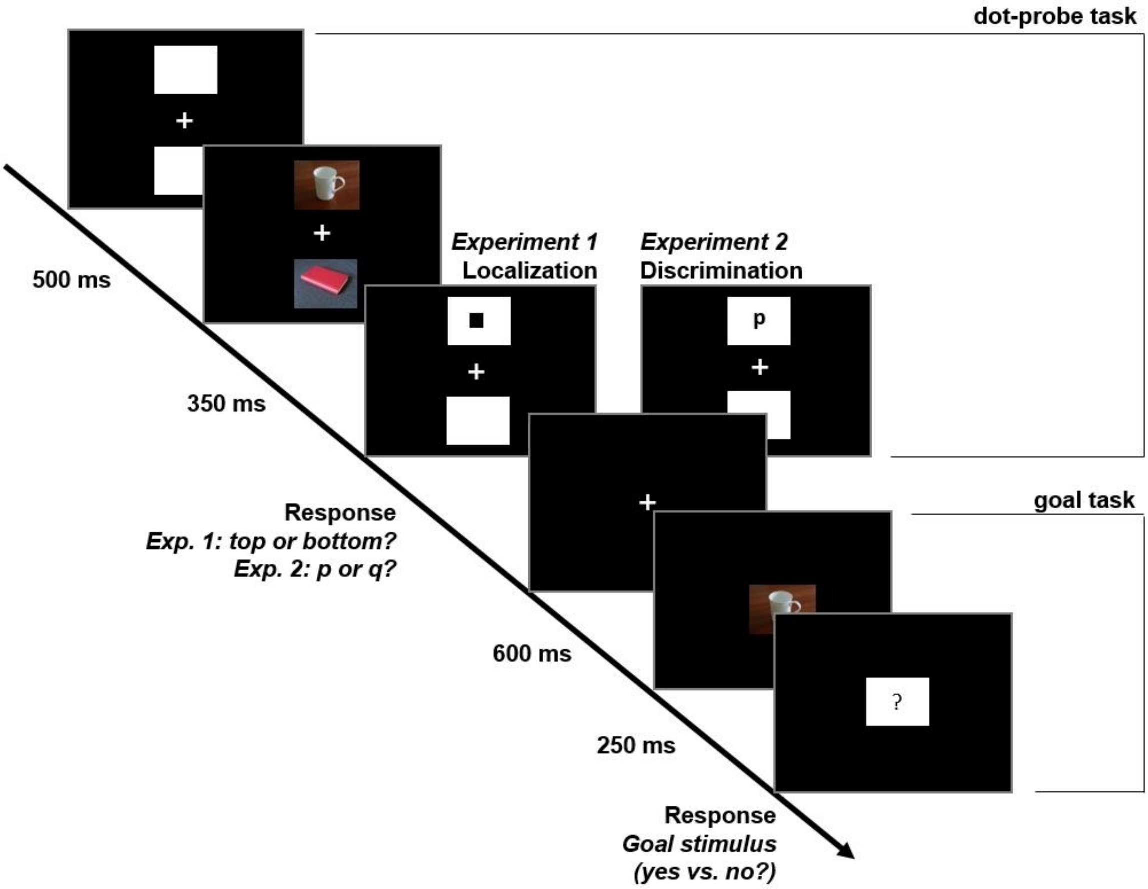
Attentional priority for temporary goals:

Evidence with a confound-free dot-probe task

A long-standing theme at the intersection of motivation and attention research is the hypothesis that stimuli associated with individuals' current goals attract attention (e.g., Klinger, 1975). In the broad sense of everyday psychology, this is certainly true. For example, if I plan to buy an electric car, I will probably be more aware of the Tesla vehicles driving by me. When I am particularly afraid of infection because I am planning to go on a vacation in the next few days, I am particularly struck by coughing people on the bus. However, can we find evidence for this hypothesis at the level of fast, involuntary basic attentional processes?

Vogt et al. (2013) explored the role of temporary goal relevance in a simple and elegant paradigm. They induced goal relevance of a stimulus by instructing participants to press a key whenever this specific picture (hereafter: the goal picture) was presented at the center of the screen (and to withhold the keypress for all other stimuli; see Figure 1). These goal trials were interspersed with a task to assess the potential attentional prioritization of the goal picture. More specifically, the authors used a variant of the well-known dot-probe task (first introduced by MacLeod et al., 1986) to assess attentional biases. In a typical trial of the dot-probe task, participants have to respond to a target stimulus presented in either of two potential target locations (e.g., above or below the center of the screen, which has to be fixated). The target display is preceded (with short stimulus onset asynchrony) by a cue display, containing stimuli at both locations. The two stimuli have a well-defined difference; in the variant by Vogt et al., one stimulus is goal-related (i.e., the goal picture) whereas the other one is motivationally neutral (see Figure 1). If participants respond more quickly to targets presented in the location of the goal picture than to targets presented in place of the neutral picture (a pattern referred to as *cueing effect*), it is assumed that attention was captured by the goal picture.

Figure 1. Schematic illustration of a typical trial of the paradigm assessing attentional priority for temporal goals (Vogt et al., 2013). In the goal-task, participants had to press the space key if the (pre-defined) goal stimulus was presented. In the dot-probe task, participants had to categorize either the location of the target (top vs. bottom; location task; Vogt et al., 2013; present Experiment 1) or to categorize the identity of the target (p or q; discrimination task; present Experiment 2). Assuming that the picture of a mug was the goal picture, a valid dot-probe trial is depicted here as the target is presented in the same location as the goal picture. For the sake of visibility, proportions are not true to scale. For copyright reasons, the pictures depicted here are not the ones actually used in the original experiment.



While the dot-probe task can be used to assess attentional bias towards motivationally relevant stimuli, it is an even more established paradigm to assess attentional bias towards emotional (but not per se motivationally relevant) stimuli. In this regard, the dot-probe task is predominantly used in two domains: First, it is used in experimental psychopathology to explore differences in attention to threat-related stimuli in anxious versus non-anxious participants (for reviews and meta-analyses see Bar-Haim et al., 2007; Frewen et al., 2008;

Kruijt et al., 2019). Second, basic cognitive research on the processing of emotional stimuli in the general population employs the dot-probe task as well to study whether negative and/or positive stimuli do generally attract attention (e.g., Brosch et al., 2008; Cooper & Langton, 2006; Holmes et al., 2005; Puls & Rothermund, 2017; Wirth & Wentura, 2018a, 2019, 2020).

Thus, Vogt and colleagues (2013; Experiment 1) conducted a dot-probe experiment in order to investigate whether threat-related or goal-relevant stimuli win the race for attention when directly competing with each other. To this end, they employed three critical cue conditions: The cue display contrasted the goal picture either with a neutral stimulus or with a threat-related stimulus; finally, the threat-related stimulus was contrasted with a neutral stimulus. Thus, they not only investigated whether the goal picture captures attention but additionally investigated whether a potential capture effect by the goal picture overrules a potential capture effect by a threat-related picture. The result was clear-cut: The goal picture caused cueing effects, either in competition with the neutral stimulus or in competition with the threat stimulus.

Vogt and colleagues (2013) replicated this basic effect of goal prioritization over threat in two other experiments. Moreover, this result was replicated in other studies with different stimulus materials (Vogt, De Houwer, & Crombez, 2011), different cue-display configurations (Vogt et al., 2010; Vogt, De Houwer, & Moors, 2011), more abstract definitions of the goal pictures (Vogt et al., 2022), shorter cue-target asynchronies (Forrest et al., 2022), and with stimuli that gained motivational relevance due to situational or individual factors – rather than task instructions (Vogt et al., 2017, Vogt, Lozo, et al., 2011, and recently Fournier & Koenig, 2023).

The location task

In all the studies cited above, however, there was a theoretical ambiguity in the interpretation of the found cueing effects because participants were asked to localize the target. That is, they were asked to indicate as accurately and as quickly as possible whether the target is presented above or below the central fixation cross (location task). Given this choice, faster responses to targets replacing the goal picture than to targets replacing the threatening picture might not have been caused by effects of spatial attention but by response-priming effects.

To explain this point in detail, let us take a step back. The dot-probe task is structurally equivalent to the exogenous cueing task (Jonides, 1981; Posner et al., 1980), used in basic attention research (see below for more details on the equivalence). In the exogenous cueing task, a cue is briefly flashed at one of two potential locations, followed by a target, presented either at the same location as the cue (valid cue) or at the opposite location (invalid cue). Participants' task is either to affirm detection of the target (in this case a portion of so-called "catch trials" includes no targets) or the target has to be categorized according to a feature that varies orthogonally to location (and validity), for example, whether the target is the letter "p" or "q". Henceforth, we will refer to this task as "discrimination task". Typical effects – that is, faster responses in valid compared to invalid trials – are interpreted as capture of spatial attention.

As mentioned above, the dot-probe paradigm is structurally equivalent to the exogenous cueing task. In both paradigms, the *difference* of the two cue locations affects spatial attention: In the exogenous cueing task, the difference is "abrupt onset cue present" (at one location) versus "abrupt onset cue absent" (at the other location); in the dot probe task (of Vogt and colleagues, 2013) the difference is "goal picture present" (at one location) versus "non-goal (i.e., neutral or threat) stimulus present" (at the other location).

Importantly, there is a good reason why the basic exogenous spatial cueing paradigm uses the discrimination task rather than the location task. The latter task could be criticized for confounding attentional processes with response-priming processes: In response priming paradigms, a target has to be categorized with regard to a (typically) binary feature (Schmidt et al., 2011; Wentura & Degner, 2010). A briefly flashed and task-irrelevant prime stimulus precedes the target. This prime stimulus can also be classified according to the target response categories. Typically, a response priming effect occurs, that is, responses are faster if the prime response category is compatible with the target-related response category than if both response categories are incompatible.

It is obvious that an exogenous cueing experiment with the location task can be reinterpreted as response priming, since the location of the cue matches the response required for the target on valid trials, but not on invalid trials. Of course, one might argue that still the dominant process is attentional capture by the cue and that the cueing effect (i.e., faster responses if the target is presented at the same location as the cue) is mainly based on the capture process. According to that argument, if (and only if) attention is allocated to the cue, the response-related feature (i.e., location) is processed and therefore boosts the cueing effect, because in the valid condition the cue-related response tendency always facilitates the target response whereas in the invalid condition the cue-related response tendency always interferes with the target response. Thus, in this case, the observed cueing effect is partially caused by response priming.

However, it is not necessary to assume a shift of spatial attention since even the pre-attentive processing of the cue might be sufficient to extract the location feature. In this case, the observed cueing effect is nothing else than a response priming effect. Evidence for this interpretation can be found in Wentura et al. (2024, Experiment 2), who briefly presented a single cue to the left or right of fixation, followed by a target stimulus, also presented to the left or right of fixation, orthogonal to the cue position. As expected, a cueing effect (i.e., faster responses on valid than on invalid trials) was found. As mentioned above, this effect can be explained by (a) a shift of attention to the cued location, (b) by response priming, or (c) by a mixture of both processes. In this experiment a strong case was made in favor of the response priming explanation: A behavioral signature of response priming processes is the congruency sequence effect (CSE), that refers to reduced response priming effects on those trials that are preceded by incongruent trials. The CSE is known for all response interference paradigms (of which response priming is a subtype) such as the Stroop task (Stroop, 1935), the flanker task (Eriksen & Eriksen, 1974) or the Simon task (Simon, 1990).[1] Experiment 2 by Wentura et al. was characterized by a pronounced CSE: after a congruent/valid trial, the response priming effect was very large; however, it disappeared completely when the preceding trial was incongruent/invalid. This latter finding makes it unlikely that spatial attention contributed to the overall effect, as this should have resulted in a residual cue validity effect when the preceding trial was incongruent.

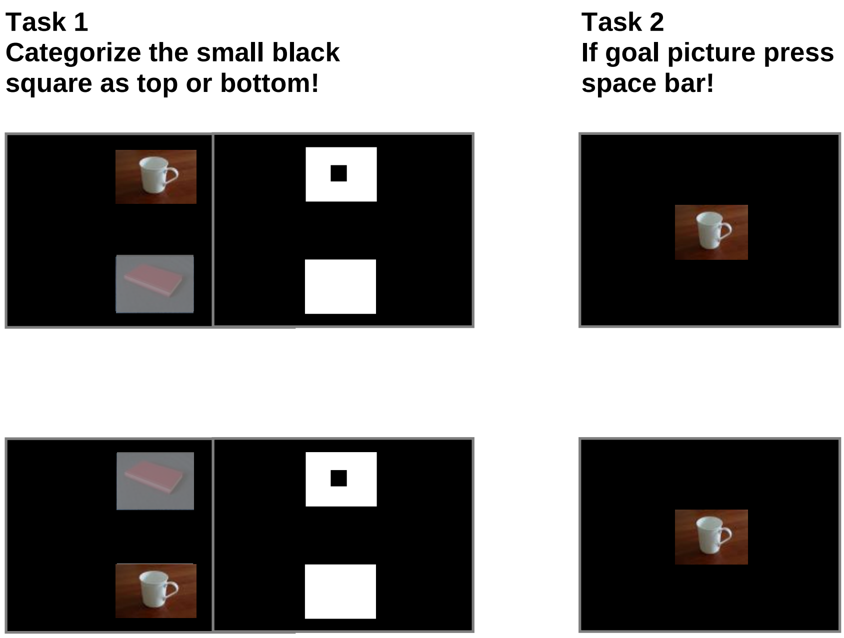
Due to the structural equivalence of dot-probe task and exogenous cueing, the argument against the location task can in principle be applied to the dot-probe task as well. As explained above for exogenous cueing, we might still argue that the dominant process is a capture of attention by the goal-related stimulus and that this process is mainly responsible for the cueing effect. In this case, a response-priming component is a potential confound that boosts the cueing effect: If attention is allocated to the goal picture, the location feature is processed and either triggers the correct (in the valid condition) or incorrect (in the invalid condition) response.

Nevertheless, as for exogenous cueing, we could interpret the situation in a way that does not refer to spatial attention processes. However, the argument is slightly more complex for the dot-probe task. Participants are faced with a double task situation, somewhat comparable to the task-switching paradigm (Kiesel et al., 2010; Monsell, 2003). In dot-probe trials, the small black rectangle has to be categorized according to location (Task 1); in goal task trials, the goal picture has to be detected (Task 2). Across tasks, only two of all potential stimuli are response relevant: the target stimulus (i.e., the small black rectangle) in the dot-

[1] There is a debate (see Duthoo et al., 2014, for a review) about the correct explanation, one of which is temporarily increased control effort (Botvinick et al., 2001). There is no need to summarize this debate here, since no explanation is based on spatial attentional processes.

probe task and the goal stimulus in the goal task; all other stimuli that will be presented can be considered distractors (i.e., they are completely task-irrelevant). Consequently, dot-probe trials that contain a goal-related cue constitute a “bi-valent” stimulus situation for participants. That is, both response-relevant stimuli – that is, the target and the goal picture – are presented. A potential perseverance of the goal task (“Detect the goal picture”) might lead to an involuntary triggering of a response. Of course, this might be a response tendency according to the actual goal task (“Detect ….and press spacebar!”). Such a tendency would be neutral with regard to the validity factor. However, it is conceivable that the response tendency is a “top!” or “bottom!” since the response set of the dot-probe task (“press a top or bottom key according to the position of the target.”) can be easily applied to the goal picture as well. In this case a potential cueing effect is actually a response priming effect. Note, in this scenario, spatial attention processes are not explicitly involved. Figure 2 illustrates this argument by showing the (potentially) subjective view of participants (hence the distractor stimulus is shown in low contrast).

Figure 2. Illustration of the response priming argument (see text for explanation).



Of course, we agree that this scenario still contains the basic idea of Vogt and colleagues (2013) of a prioritization of goal related stimuli. However, the reframing in terms of task-switching and response priming effects sheds a different, less spectacular light on the cueing effect.

We can easily remove this confound by changing the location task to the discrimination task (i.e., instructing participants to categorize the target as “p” or “q”). Now, a spatial attention scenario and a response-priming scenario diverge in regards to behavioral outcomes. If prioritization of goal-related stimuli consists only in an asymmetrical processing of cue displays (i.e., a goal-related stimulus triggers response-related processes), the cueing effect should vanish because the task set (“press key *x* for “p” and key *y* for “q”) cannot be applied to the goal picture. If, however, prioritization of goal-related stimuli consists in a capture of spatial attention, this should be observable in a non-confounded cueing effect.

Therefore, our study aims to investigate whether the effects of attentional priority for temporary goals can be replicated with a non-confounded target-response task, namely target discrimination. To this end, with Experiment 1 we replicated Vogt et al.’s (2013) Experiment 1 – that is, a dot-probe procedure with the location task – to establish a kind of baseline. In Experiment 2, we repeated exactly the same procedure with the exception of changing the target response to discrimination. That is, the target on a given trial was either the letter “p” or the letter “q” and participants had to identify the letter. We recruited participants for Experiments 1 and 2 in parallel with random assignment. This allows for unambiguous across-experiments analyses that will be presented subsequently. Similar to the original experiment, there were three basic cue display types in our experiments: (1) Goal picture vs. neutral picture; (2) goal picture vs. threatening picture; (3) threatening picture vs. neutral picture. If the results obtained by Vogt et al. (2013) were indeed caused by processes of spatial attention, we should be able to find a significant (positive) cueing effect in conditions

(1) and (2) both in Experiment 1 (target-location task) and in Experiment 2 (target-discrimination task). If the results were completely caused by response-priming effects, we should only find this cueing effect in Experiment 1 but not in Experiment 2. It is also possible that both processes were at play in the experiment by Vogt and colleagues (2013). If this is the case, it is possible that the cueing effect found in Experiment 1 significantly exceeds the one in Experiment 2.

Experiment 1

Method

Transparency and openness

We report how we determined our sample size, any data exclusions, all manipulations, and all measures in this study. The data and the program code for all experiments are available on the Open Science Framework (OSF). These files can be accessed via the following link: https://osf.io/z52hc/?view_only=732b43f7534d4f7f937f39e50c5e098a (Wentura & Wirth, 2023, August 16). We pre-registered our experiments on aspredicted.org. The preregistration can be accessed via the following link: https://aspredicted.org/SW4_2LS There were some deviations from the preregistration. We listed and justified them in Appendix B. The study was approved by the Ethics Committee of the Faculty of Human and Business Sciences, Saarland University, Saarbrücken, Germany.

Participants

The final sample consisted of 35 participants (10 men, 25 women). Age ranged from 19 to 30 years (Md = 22 years). Participants were students at the University of Saarland, Germany. (Ethnicity was not coded explicitly. The vast majority can be assumed to have a White ethnic background.)

The cueing effects in dot-probe trials where (emotionally neutral) goal pictures competed with (a) neutral pictures and (b) threatening pictures in Vogt et al.'s (2013; Experiment 1) study were of size $d_Z = 1.46$ and $d_Z = 0.75$, respectively. In order to be able to detect the smaller effect with a power of 1-β = .8 (given α = .05) the required sample size would be N = 16 (G*Power; Faul et al., 2007). However, in order to be able to detect differences in cueing effects between the two experiments with a size of $d_S = 0.7$, we needed to test n = 34 participants for each experiment (for power 1-β = .8).[2] In order to compensate for potential exclusions of participants, we tested a few more participants than the required total number of N = 34.

Thus, we recruited N= 39 participants.[3] Data of n = 4 participants had to be discarded according to our preregistered exclusion criteria because accuracy rates in either the dot-probe task or the goal-picture task was more than three inter-quartile ranges below the first quartile of the distribution of all participants. There were no outliers with regard to mean response times. With N = 35 we were able to detect effects of size $d_Z = .75$ (i.e., the smaller one of Vogt et al.'s effects) within this group with a power of 1-β = .99.

Design

We employed a 3 (*trial type:* goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: invalid vs. valid cue) within-participants design. Validity is always defined for the category mentioned first in the trial type conditions. In accordance with Vogt and colleagues (2013; Exp. 1) we added a fourth trial type condition "threatening picture/ threatening picture" to balance the presentation of a priori affectively neutral (i.e., goal and neutral pictures) and threatening pictures. Of course, validity cannot be meaningfully defined for these filler trials.

Materials

---

[2] A between-participants effect size of $d_S = 0.7$ results, for example, if $d_Z = .75$ holds for the location sample whereas the effect is almost null (i.e., $d_Z = .05$) for the discrimination sample (if standard deviations of the cueing differences are equal). Thus, we wanted to be prepared for the case that the discrimination sample yield a null effect.

[3] For unknown technical reasons, responses were not recorded in the data sets of two additional participants. These sets could not be used for any further analyses.

We used the same pictures as Vogt et al. (2013) for the main phase of the experiment, all taken from the International Affective Picture System (Lang et al., 2008; see *Appendix A*). There were two sets of five stimuli each, two neutral pictures and two threat-related pictures for the dot-probe and goal tasks plus one additional neutral picture for the goal task. Each set contained one picture corresponding to evolutionary threat (attacking snake, barking pit bull) and one picture containing a human attack scene (knife attack, gun attack). Participants were assigned to either Set A or B; participants were assigned either neutral stimulus 1 or 2 (of the respective set) as the goal-relevant stimulus. For the practice trials, we selected a further set of IAPS pictures (see *Appendix A*).

Procedure

The study was conducted on four PCs equipped with 17" CRT monitors using a resolution of 1,024 × 768 Pixels, a refresh rate of 100 Hz, and a color depth of 32 bit. Eye-screen distance was approximately 60 cm. The experimental routine was programmed using PsychoPy (Peirce et al., 2022).

Dot-probe task. A trial of the dot-probe task started with the presentation of a fixation cross at the center of the screen and two rectangles (4.6 cm × 6.1 cm) above and below the fixation cross (see Figure 1). The centers of the rectangles had a distance of 4.6 cm from the center of the screen. After 500 ms, the rectangles were filled by two cue pictures for 350 ms.[4] Subsequently, the target appeared in one of the rectangles for 850 ms or until participant response.[5] The target was a small black rectangle (0.5 × 0.5 cm). Participants had to

[4] That is, the cue-target asynchrony (CTA) was 350 ms, as introduced by Vogt and colleagues (2013). This is not untypical for dot-probe studies; however it is a rather long CTA if compared to basic exogenous cueing experiments. Therefore, we used a CTA of 100 ms in our previous research (Wirth & Wentura, 2018a, 2019, 2020) to more plausibly address early capture processes and to avoid distortion of effects by eye movements (Petrova et al., 2013). In the present experiment, however, the replication character of the study was more important, so we kept the CTA of 350 ms.

[5] This is a minor unintended (i.e., due to a program error) deviance from Vogt et al. (2013) who used a maximal duration of 1500 ms instead of 850 ms. However, distribution of raw

categorize the position of the target (top vs. bottom) by pressing either the “4” or the “5” key on the num pad (assignment counterbalanced) with their right index or middle finger, respectively. Participants were asked to maintain attention at the fixation cross and to respond as quickly and as accurately as possible to the target. There were three blocks of 64 trials each. Within each block, each of the four types of dot-probe trials (i.e., goal versus neutral, goal versus threat, threat versus neutral, and threat versus threat) was presented 16 times. This number of trials results from the complete combination of the factors target location × cue validity × cue identity × target identity. Cue identity is relevant only for threat/neutral and goal/threat trial types because there were two different threat-related stimuli for each participant (but only one goal-related and one neutral cue). Target identity has no meaning in the present experiment since there was only one target in the location task (i.e., the small black square, see above). The factor was implemented in anticipation of Experiment 2; there it refers to whether “p” or “q” is presented as the target. In accordance with Vogt et al. (2013) we implemented the trial type threat vs. threat in order to balance the appearance of affectively neutral (i.e., the goal and the neutral picture) and threatening pictures. For these filler trials, the assignment of the two threat-related stimuli to the “validity” condition was balanced. Thus, any difference between the “valid” and “invalid” condition is purely coincidental.

Goal task. A dot-probe trial was always followed by a goal-task trial. Following a delay of 600 ms, one out of five pictures was presented at the center of the screen for 250 ms, followed by a question mark for 2000 ms or until participant response. Participants were instructed to press the space bar with their left hand if the stimulus was the goal picture. They received feedback in each trial, that is, either “*richtig*” (“correct”), written in green, or

---

RTs shows that the cut at 850 ms had only marginal effects (e.g., the vast majority of participants had an individual outlier criterion, see *Results*, that was below 850 ms). Note that in a comparable study, Koster et al. (2007) used 750 ms as a general outlier criterion.

"*falsch*" ("wrong"), written in red, appeared on the screen for 200 ms. Each of the four pictures used in the dot-probe task was presented in 12 trials of a block. As in Vogt et al. (2013), in a quarter of all goal task trials a fifth (neutral) stimulus was presented. The inter-trial interval was 600 ms.

Before the main trials, participants exercised the tasks with 20 practice trials. In the practice and main phases, different stimuli were used. The main experiment comprised 192 trials (i.e., three blocks of 64 trials) and lasted about 30 minutes. A self-paced break was taken after each block of 64 trials. Participants received feedback about their performance in the goal task after each block.

Results

Average dot-probe accuracy was M = 98.5% (SD = 1.7%). For the response time (RT) analysis, RTs below 150 ms were excluded, as were RTs more than 1.5 interquartile ranges above the third quartile of the individual participant's distribution (Tukey, 1977). This led to the exclusion of 5.15 % of all trials with correct responses. Table 1 shows average RTs and accuracy rates as a function of the experimental factors.

Analyses of RTs

A 3 (*trial type*: goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: valid vs. invalid) MANOVA for repeated measures (O'Brien & Kaiser, 1985) yielded significant main effects for trial type, $F(2, 33) = 4.38$, $p = .021$, $\eta_p^2 = .210$, and for cue validity, $F(1, 34) = 75.15$, $p < .001$, $\eta_p^2 = .688$.. The main effects were further qualified by a significant interaction, $F(2, 33) = 23.37$, $p < .001$, $\eta_p^2 = .586$.

This interaction is almost completely caused by the first Helmert contrast (i.e., threat/neutral vs. all trials involving goal pictures), $F(1, 34) = 47.86$, $p < .001$, $\eta_p^2 = .585$, whereas the second contrast (i.e., goal/neutral vs. goal/threat trials) was not significant, $F(1, 34) = 3.10$, $p = .088$, $\eta_p^2 = .083$.The two conditions involving goal pictures were associated with strong cueing effects, $t(34) = 10.80$, $p < .001$, $d_Z = 1.82$, for the effect collapsed across goal/neutral and goal/threat trials. There was, however, no cueing effect in the threat/neutral condition, $t(34) = 0.31$, $p = .756$, $d_Z = 0.05$.

Table 1

Mean RTs and Cueing Scores (in ms) as a Function of Cue Validity, Cue Type, and Experiment (Task)

Task	Cue validity		
Trial type	Valid	Invalid	Cueing-Score
Experiment 1 (Location Task)			
Goal vs. Neutral	439 (96.5%)	478 (93.7%)	38 [5]
Goal vs. Threat	442 (96.7%)	490 (93.3%)	47 [5]
Threat vs. Neutral	463 (96.0%)	464 (96.5%)	2 [5]
Threat vs. Threat (Filler trials)	*467 (97.0%)*		
Experiment 2 (Discrimination Task)			
Goal vs. Neutral	566 (91.3%)	586 (89.6%)	20 [6]
Goal vs. Threat	570 (93.5%)	596 (89.3%)	26 [8]
Threat vs. Neutral	582 (92.5%)	572 (91.0%)	-10 [5]
Threat vs. Threat (Filler trials)	*582 (92.0%)*		

Note. Accuracy rates are given in parentheses. Cueing score = $RT_{invalid} - RT_{valid}$; SEs are given in brackets. Discrepancies between mean RTs and cueing scores are due to rounding. "Valid" refers to trials in which the target replaced the cue category first mentioned under trial type; "invalid" refers to trials in which the target replaced the cue category mentioned second under trial type. For the filler trials (i.e., threat vs. threat), the distinction of "valid" versus "invalid" is meaningless.

Analyses of error rates

For error rates, a 3 (*trial type*: goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: valid vs. invalid) MANOVA for repeated measures (O'Brien & Kaiser, 1985) yielded a non-significant main effect of trial type, $F(2, 33) = 1.85$, $p = .173$, $\eta_p^2 = .101$, and a significant main effect of cue validity, $F(1, 34) = 10.32$, $p = .003$, $\eta_p^2 = .233$, that was further qualified by a significant interaction, $F(2, 33) = 5.28$, $p = .010$, $\eta_p^2 = .242$.

Again, the interaction is almost completely caused by the first Helmert contrast (i.e., threat/neutral vs. all trials involving goal pictures), $F(1, 34) = 10.85$, $p = .002$, $\eta_p^2 = .242$ whereas the second contrast (i.e., goal/neutral vs. goal/threat trials) was non-significant, $F < 1$.. There was no cueing effect in the threat/neutral condition, $t(34) = 0.87$, $p = .392$, $d_Z = 0.15$, whereas the two conditions involving goal pictures were associated with substantial cueing effects, $t(34) = 3.73$, $p < .001$, $d_Z = 0.63$, for the effect collapsed across goal/neutral and goal/threat trials.

Discussion

The results of Vogt et al. (2013; Experiment 1) were almost perfectly replicated when a location task was used. There were large cueing effects for goal-related stimuli, but no effect for threat-related ones. However, as explained in the introduction section, the effect cannot be unambigously interpreted in terms of attentional processes. Therefore, our main experiment (Experiment 2) is a replication using a target discrimination task. With this change, any cueing effects can no longer be (alternatively) interpreted as response-related processes and can only be be interpreted in terms of genuine attentional processes.

Experiment 2

Method

Transparency and openness

See the corresponding section of Experiment 1.

Participants

The final sample consisted of 30 participants (8 men, 22 women). Age ranged from 19 to 27 years (Md = 22 years). Participants were students at the University of Saarland, Germany. (Ethnicity was not coded explicitly. The vast majority can be assumed to have a white ethnic background.) The power planning was already explained in the *Participants* section of Experiment 1. That is, we initially aimed for $N = 34$ participants. In order to compensate for potential exclusions of participants, we recruited $n = 39$.[6] Data of $n = 8$ participants had to be discarded according to our preregistered exclusion criteria because accuracy rates in either the dot-probe task or the goal-picture task was more than three inter-quartile ranges below the first quartile of the distribution of all participants. There were no outliers with regard to mean response times. Thus, sample size in the discrimination task sample was somewhat lower than planned. However, with N=30 we were able to detect effects of size $d_Z = .75$ (i.e., the smaller one of Vogt et al.'s effects) within this group with a power of $1\text{-}\beta = .98$.

Design, Materials, and Procedure

Design, Materials, and Procedure were exactly as in Experiment 1, except for the target stimuli and the target task within the dot-probe trials. Now, the target was either the letter “p” or “q” (app. 0.5 × 1.0 cm). Participants had to categorize the identity by pressing either the “4” or the “5” key on the number pad (assignment counterbalanced) with their right index or middle finger, respectively. Participants were asked to maintain attention at the fixation cross and to respond as quickly and as accurately as possible to the target.

Results

Average dot-probe accuracy was $M = 96.4$ % ($SD = 2.8$%). For the response time (RT) analysis, RTs below 150 ms were excluded, as were RTs more than 1.5 interquartile ranges

[6] For unknown technical reasons, responses were not recorded in the data sets of four additional participants. These sets could not be used for any further analyses.

above the third quartile of the individual participant's distribution (Tukey, 1977). This led to the exclusion of 2.44 % of all trials with correct responses. Table 1 shows average RTs and accuracy rates as a function of the experimental factors.

Analyses of RTs

A 3 (*trial type*: goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: valid vs. invalid) MANOVA for repeated measures (O'Brien & Kaiser, 1985) yielded a non-significant effect of trial type, $F(2, 28) = 2.62$, $p = .091$, $\eta_p^2 = .157$, and a significant main effect of cue validity, $F(1, 29) = 6.73$, $p = .015$, $\eta_p^2 = .188$ that were further qualified by a significant interaction, $F(2, 28) = 8.31$, $p = .001$, $\eta_p^2 = .372$.

This interaction is almost completely caused by the first Helmert contrast (i.e., threat/neutral vs. all trials involving goal pictures), $F(1, 29) = 17.09$, $p < .001$, $\eta_p^2 = .371$ whereas the second contrast (i.e., goal/neutral vs goal/threat trials) was not significant, $F < 1$. As in Experiment 1, the two conditions involving goal pictures were associated with strong cueing effects, $t(29) = 3.69$, $p < .001$, $d_Z = 0.67$, for the effect collapsed across goal/neutral and goal/threat trials. For the threat/neutral condition, the cueing effect had a negative sign and can be considered significant,[7] $z = 2.03$, $p = .043$, $d_Z = 0.35$.

Analyses of error rates

For error rates, a 3 (*trial type*: goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: valid vs. invalid) MANOVA for repeated measures (O'Brien & Kaiser, 1985) yielded a significant main effect for cue validity, $F(1, 29) = 4.89$, $p = .035$, $\eta_p^2 = .144$, but not for trial type, $F(2,28) = 1.05$, $p = .365$, $\eta_p^2 = .070$, and no significant interaction, $F(2, 28) = 1.51$, $p = .238$, $\eta_p^2 = .097$.

[7] Wilcoxon test (due to non-normality of the cueing score; Shapiro-Wilks .94, $p = .089$; because of the rather small *N*, we adopted an alpha level of .10 to reduce the risk of a Type II error).

The overall cueing effect is in the expected direction, that is, more errors occurred in the invalid trials compared to the valid ones. Thus, for the goal-related cueing effects, RT and error effects are in line. For the threat/neutral condition, a substantial positive cueing effect for errors would have indicated a speed-accuracy tradeoff. However, the cueing effect is small and non-significant, $t(29) = 1.17$, $p = .250$, $d_Z = 0.21$.

Discussion

Experiment 2 yielded a clear result: The effect of attentional priority for temporal goals can be found in a non-confounded design as well. The pattern of effects closely mimics the one for the location task. However, there are two notable differences: first, the goal-related effects are smaller in Experiment 2 than in Experiment 1; second, whereas the threat-related effect was null in Experiment 1, in Experiment 2 a small negative effect was obtained.

Cross-Experiments Analyses

As already noted in the *Participants* section of Experiment 1, participants were assigned to Experiment 1 and 2 on a random basis. Thus, we can present a cross-experiments analysis with significance tests that are not subject to any reservations. For the remainder of this cross-experiments analysis we will use the label *task* with the conditions *location* (i.e., Experiment 1) and *discrimination* (i.e., Experiment 2).

Analyses of RTs

A 2 (*task*: location [Exp. 1] vs. discrimination [Exp. 2]) × 3 (*trial type*: goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: valid vs. invalid) mixed-design MANOVA for repeated measures yielded a main effect of task, $F(1, 63) = 63.77$, $p < .001$, $\eta_p^2 = .503$; not surprisingly, the discrimination task produced somewhat slower responses. There was a main effect of trial type, $F(2, 62) = 6.35$, $p = .003$, $\eta_p^2 = .170$. The first Helmert contrast (threat/neutral trials vs. all trials involving goal pictures, $F < 1$, and the second Helmert contrast (goal/neutral vs goal/threat trials), $F(1, 63) = 12.92$, $p < .001$, $\eta_p^2 = .170$, showed that RTs were increased for goal/threat trials (see Table 1).

There were three further significant effects, all involving the cue validity factor. There was a main effect of cue validity, $F(1, 63) = 54.19$, $p < .001$, $\eta_p^2 = .462$, which was moderated by trial type, $F(2, 62) = 28.50$, $p < .001$, $\eta_p^2 = .479$, and task, $F(1, 63) = 9.68$, $p = .003$, $\eta_p^2 = .133$. The trial type × cue validity interaction is almost completely caused by the first Helmert contrast (i.e., threat/neutral vs. all trials involving goal pictures), $F(1, 63) = 57.48$, $p < .001$, $\eta_p^2 = .477$, whereas the second contrast was not significant, $F(1, 63) = 3.06$, $p = .085$, $\eta_p^2 = .046$. Across tasks, there was no cueing effect in the threat/neutral condition, $t(64) = 1.02$, $p = .313$, $d_Z = 0.13$, whereas the two conditions involving goal pictures were associated with strong cueing effects, $t(64) = 8.96$, $p < .001$, $d_Z = 1.11$, for the effect collapsed across goal/neutral and goal/threat trials. The two remaining tests – that is, trial type × task and trial type × cue validity × task – were not significant, both *F*s < 1.

The task × cue validity interaction indicates that cueing scores are overall smaller in the discrimination condition. As planned (see preregistration), we report t-tests for the conditions with goal-related cues. For goal/neutral trials, the difference in mean cueing scores is significant, $t(63) = 2.44$, $p = .017$, $d_S = 0.61$. However, cueing scores for both tasks were significant, $t(34) = 8.41$, $p < .001$, $d_Z = 1.42$, and $t(29) = 3.09$, $p = .004$, $d_Z = 0.56$ for the location and discrimination task, respectively. For goal/threat trials, the difference in mean cueing scores is significant as well, $t(49.14) = 2.34$, $p = .023$, $d_S = 0.60$. Again, cueing scores for both tasks were significant, $t(34) = 9.83$, $p < .001$, $d_Z = 1.66$, and $t(29) = 3.31$, $p = .003$, $d_Z = 0.60$ for the location and discrimination task, respectively.

Not pre-registered, but self-suggesting is a control analysis that reveals whether the difference in goal cueing effects is a mere by-product of the difference in mean RTs between the two samples. Thus, we regressed the cueing scores on task and mean RT as predictors. For both goal/neutral trials and goal/threat trials, the difference in mean cueing scores remains significant if mean RT is introduced as a competing predictor to task, $t(62) = 2.13$, $p = .037$ and $t(62) = 2.65$, $p = .010$, respectively. Mean RT was not a significant predictor in these regressions, both $|t|$s < 1.33.

Threat/neutral trials were not the focus of our hypotheses. Exploratorily, we analyzed them as well. We already reported a null effect in the *Results* section of Experiment 1 and a negatively signed effect for Experiment 2. However, the difference in mean cueing scores is not significant, $z = 1.68$, $p = .092$, $d_S = 0.39$.[8]

Analyses of error rates

For error rates, a 2 (task) × 3 (trial type) × 2 (cue validity) mixed-design MANOVA for repeated measures yielded a significant main effect for task, $F(1, 63) = 7.54$, $p = .008$, $\eta_p^2 = .107$, with the discrimination task producing somewhat more errors, a significant main effect of cue validity, $F(1, 63) = 12.89$, $p < .001$, $\eta_p^2 = .170$, and a significant trial type × cue validity interaction, $F(2,62) = 4.21$, $p = .019$, $\eta_p^2 = .120$. The interaction is almost completely caused by the first Helmert contrast, $F(1, 63) = 4.83$, $p = .032$, $\eta_p^2 = .071$ whereas the second contrast is not significant, $F(1, 63) = 1.16$, $p = .286$, $\eta_p^2 = .018$. That is, there is no cueing effect in the threat/neutral condition, $t(64) = 0.54$, $p = .591$, $d_Z = 0.07$, whereas the two goal conditions are associated with cueing effects, $t(64) = 3.55$, $p < .001$, $d_Z = 0.44$, for the effect collapsed across goal/neutral and goal/threat. All other tests of the overall analysis are not significant, all Fs < 1 except $F(2,62) = 2.43$, $p = .092$, $\eta_p^2 = .037$ for the main effect of trial type.

It is sometimes recommended to integrate RT and accuracy data into one measure, the Inverse Efficiency Scores (IES; Townsend & Ashby, 1978). IES are defined by mean RTs divided by the proportion of correct responses. However, Bruyer and Brysbaert (2011) can be read as a note of caution regarding the use of IES. Nevertheless, we present analyses with IES

[8] Mann-Whitney-U test due to non-normality of the cueing score within the discrimination task sample.

in *Appendix C*. They can be taken as a further piece of information to evaluate the comparison of the location and discrimination task. In general, the results using IES were somewhat weaker than the RT analyses (due to increased variances, see Bruyer and Brysbaert, and the greater likelihood of outlier values). However, there was one detail that might deserve further discussion. While the RT results for the cueing effects of goal vs. neutral trials were essentially replicated in the IES analyses, a divergence was found for the goal vs. threat trials: IES cueing effects gathered with the discrimination task were now at approximately the same level as IES cueing effects gathered with the location task.

General Discussion

The present study yielded several noteworthy results: First, the results of Vogt et al. (2013; Experiment 1) were almost perfectly replicated when a location task was used (Experiment 1). There were large cueing effects for goal-related stimuli, but no effect for threat-related ones. Second, we found almost the same result pattern with a discrimination task, that is, there were goal-related cueing effects (Experiment 2). However, effects were halved with regard to absolute means and more than halved with regard to effect size. Third, with the discrimination task, we found a *negatively* signed medium-to-small effect for threat pictures that was completely missing in the location task sample.

How can this pattern of results be interpreted? In the introduction section, we asked the question whether the effects found by Vogt et al. (2013) – which were replicated in our location sample – were in total, partially, or not at all caused by a response priming confound. The goal-related effects found with the discrimination task show that response priming is definitely not the sole driver of the results found in the location task. They corroborate the basic assumption of Vogt et al. (2013) that stimuli associated with current goals do indeed attract attention. Due to removing the confound, the effects can be unambiguously attributed to shifts of spatial attention.

However, the cueing effects are significantly reduced with the discrimination task. There are two possible causes for this pattern: Firstly, the effects found with the location task were partially caused by confounding response priming processes; this part of the effect is missing with the discrimination task. Secondly, however, we need to discuss a possibility that is suggested by the obvious observation that the two tasks differ in their demands, as can easily be seen from the increased response times and errors in the discrimination task compared to the location task. In principle, it could be the case that the basic attentional processes are the same in both tasks, but that the consequences of attentional shifts are distributed differently in response speed and errors for the two tasks. To test for this, we additionally analyzed the inverse efficiency scores (IES; see Appendix B), an index that integrates response speed and accuracy (Townsend & Ashby, 1978). We should treat these analyses with some caution, as IES tends to be noisier and more prone to outliers (Bruyer & Brysbaert, 2011). Nevertheless, for the goal/neutral condition the IES analyses essentially replicated the RT analyses: The cueing effect for the location sample clearly exceeds that for the discrimination sample. Thus, our first interpretation that the cueing effect found with the location task is somewhat inflated by response priming processes seems to be valid for the goal/neutral condition.

For the goal/threat condition, however, the IES analysis gave a slightly different picture than the RT analysis: The cueing effects found in the discrimination task were similar to those found in the location task. This could be taken as an indication that the processes in the goal/threat condition are somehow different from those found in the goal/neutral condition. This leads to the next point.

One detail of the result pattern found with the discrimination task is different from the pattern found with the location task: With the discrimination task we found a *negatively* signed medium-to-small effect for threat pictures (that are competing with neutral pictures) that was absent in the location task sample. Thus, if we consider this *negatively* signed effect as a robust result (i.e., not as a false positive result), the premise of the research of Vogt et al. (2013) – that is, that goal-related stimuli compete with threat-related ones for attentional capture – does not hold. In fact, the three cueing effects – the goal-vs-neutral effect, the goal-vs-threat effect, and the threat-vs-neutral effect – complement each other in the discrimination task sample: If we assume an attentional capture effect for goal-related stimuli and a (to-be-discussed) process causing a negative effect for threat stimuli, the cueing effect in the goal versus threat condition is increased compared to the goal versus neutral condition. This pattern was indeed found in the discrimination sample. The pattern for threat-related stimuli, however, was somewhat hidden in the location sample, probably due to the dominance of the goal-related effects that were boosted by the response-priming confound.

Why did we find a negatively signed effect for threat-related stimuli? The expectation of Vogt et al. (2013) – as well as ours – to find a positive effect should not obscure the fact that negative effects were (a) sometimes predicted theoretically, and (b) also found. With regard to (a), the seminal theories of Williams et al. (1988, 1997) and Mogg and Bradley (1998) on attentional biases in anxiety include the prediction of avoidance from mild threatening stimuli in low anxious persons. Of course, these two constraints do not map directly to the stimuli and samples used in Vogt et al. (2013) and our study. However, even if the threat stimuli chosen in these two studies were highly threatening with regard to norm values, the massive repetition of only two pictures per participant could have transformed initially highly threatening pictures into mild ones through habituation. An unselected sample (in terms of anxiety) cannot of course be equated with a low anxiety sample, but unselected samples are usually dominated by non-anxious participants.[9]

[9] For example, in previous studies (Wirth & Wentura, 2018a, 2018b, 2019) where we asked unselected samples to complete the trait scale of the Stait-Trait-Anxiety Inventory (STAI; Laux et al., 1981), sample average scores ranged from 39.5 to 40.6 (with the scale of possible values ranging from 20 to 80).

With regard to (b), indeed occasionally *negative* cueing effects were reported. We focus here on studies using picture stimuli (for negative effects with words and faces, see, e.g., Bradley et al., 1997, Cooper & Langton, 2006, Mogg et al., 1995): Yiend and Mathews (2001) found a significant negative effect for low anxious participants. Koster et al. (2005) as well as Mogg et al. (2000, Exp. 2) confirmed this result for mild threat scenes (but see Mogg et al., 2000, Exp. 1, and Mogg et al., 2004, for null results with mild threat; for positive effects, especially with high threat stimuli, see, e.g., Lipp & Derakshan, 2005; Mogg et al., 2000, Exp. 1; Onie & Most, 2022). All mentioned picture studies used a CTA of 500 ms. Long CTAs such as 500 ms and the CTA of 350 ms employed in the present study allow to explain *negative* cueing effects as a consequence of avoidance processes. Two slightly different versions of this argument exist: In the first version, an initial attention capture for threat-related pictures is assumed; however, it is followed by a disengagement of attention. (This is akin to *inhibition of return*; e.g., Klein, 2000; for a discussion in the context of the dot-probe task, see Boga & Wentura, 2024.) In the second version, it is assumed that initial attention is randomly allocated either to the threat-related cue or the neutral cue. If initial attention is allocated to the threat-related picture, it is quickly withdrawn; this is not the case for neutral cues that randomly received initial attention. Consequently, on average, one obtains a negatively signed cueing effect for threat-related pictures.[10]

Constraints on Generality and other Limitations

Our study was a behavioral study; thus, we are not able to disentangle the contribution of different attentional processes with the dependent measures employed. Therefore, it might

[10] We should additionally mention here a more sobering interpretation of the negative effect: The negative effect might be a stimulus artefact, since the comparison of "threat versus neutral" is only *quasi-experimental*. That is, especially in our (and Vogt et al.'s, 2013) small stimulus sets, there is a substantial risk that some low-level image features, whose presence or absence are only arbitrarily associated with "threat versus neutrality", affect spatial attention. However, this speculation is not very plausible because the threat stimuli behave similarly in the contrasts with neutral and goal pictures, respectively (see above).

be worthwhile to assess the attentional processes using the N2pc, an event-related potential component. The N2pc component reflects the focus of covert attention on a peripheral stimulus (for a review, see Luck, 2012; for an application within the dot-probe task, see Wirth & Wentura, 2023). An alternative might be to assess overt attentional shifts by eye-tracking (see Petrova et al., 2013).

Regarding constraints on generality, readers should keep in mind the following three points: First, the goal task in the experiment by Vogt et al. (2013) and the present one is a minimal one: Participants simply have to press a key if a specific stimulus is presented during a goal trial. Second, the assessment of threat is operationalized by only two pictures for each participant with a total of four pictures.

Conclusions

In conclusion, the present study confirms the hypothesis of Vogt and colleagues (2013) of an attentional prioritization of stimuli that are related to temporary goals with a confound-free design. The confound-free attentional bias effect (as assessed with the discrimination task) is, however, smaller than the one obtained with the location task. This result suggests that cueing effects found with the location task are not a pure measure of attention allocation, but they can be confounded with response priming processes. While in some contexts these response priming effects might only artificially increase the observed size of a cueing effect, they do have the potential to produce significant cueing effects despite the complete absence of an attentional bias. Therefore, we generally recommend to use discrimination tasks in the dot-probe paradigm.

Acknowledgments. We thank Anna Ulrike Born and Lea Haas for help with data collection.

Declaration of Interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. This research was supported by a grant from the German Research Foundation to Benedikt Wirth and Dirk Wentura (WI 5758/1-1; WE 2284/16-1).

Conflicts of interest/Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval. The study was approved by the Ethics Committee of the Faculty of Human and Business Sciences, Saarland University, Saarbrücken, Germany.

Consent to participate. Informed consent was obtained from all individual participants included in the study.

Consent for publication. Not applicable.

Availability of data and materials. The data for all experiments are available on the Open Science Framework (OSF). These files can be accessed via the following link: https://osf.io/z52hc/?view_only=732b43f7534d4f7f937f39e50c5e098a (Wentura & Wirth, 2023, August 16).

Code availability. The program code for all experiments are available on the Open Science Framework (OSF). These files can be accessed via the following link: https://osf.io/z52hc/?view_only=732b43f7534d4f7f937f39e50c5e098a (Wentura & Wirth, 2023, August 16).

Authors' contributions. Dirk Wentura and Benedikt Wirth contributed equally to conceptualization, investigation, methodology, project administration, writing–review and editing, and formal analysis. Benedikt Wirth served as lead for software and data curation; Dirk Wentura served as lead for writing–original draft.

References

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1-24. https://doi.org/10.1037/0033-2909.133.1.1

Boga, M., & Wentura, D. (2024). The assessment of attentional bias to cleanliness stimuli in different versions of the dot-probe task: Evidence for a motivational account. *Quarterly Journal of Experimental Psychology, Advance online publication*. https://doi.org/https://doi.org/10.1177/17470218241281694

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-652. https://doi.org/10.1037/0033-295X.108.3.624

Bradley, B. P., Mogg, K., & Lee, S. C. (1997). Attentional biases for negative information in induced and naturally occurring dysphoria. *Behaviour Research and Therapy*, *35*(10), 911-927. https://doi.org/10.1016/s0005-7967(97)00053-3

Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological Science*, *19*(4), 362-370. https://doi.org/10.1111/j.1467-9280.2008.02094.x

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*(1), 5-13. https://doi.org/10.5334/pb-51-1-5

Champely , S. (2018). *PairedData: Paired Data Analysis. R package version 1.1.1*. https://CRAN.R-project.org/package=PairedData

Cooper, R. M., & Langton, S. R. H. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour Research and Therapy, 44*(9), 1321-1329. https://doi.org/10.1016/j.brat.2005.10.004

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: an update. *Frontiers in Psychology, 5*, Article 1001. https://doi.org/10.3389/fpsyg.2014.01001

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143-149.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Forrest, J., Vogt, J., McDonald, C., Searle, C., & Sakaki, M. (2022). Blind to Threat: The Presence of Temporary Goals Prevents Attention to Imminent Threat Already at Early Stages of Attention Allocation. *Motivation Science, 8*(3), 239-251. https://doi.org/10.1037/mot0000261

Fournier, H., & Koenig, O. (2023). Combined effects of intrinsic and goal relevances on attention and action tendency during the emotional episode. *Emotion, 23*(2), 425-436. https://doi.org/10.1037/emo0001081

Frewen, P. A., Dozois, D. J. A., Joanisse, M. F., & Neufeld, R. W. J. (2008). Selective attention to threat versus reward: Meta-analysis and neural-network modeling of the dot-probe task. *Clinical Psychology Review, 28*(2), 307-337. https://doi.org/https://doi.org/10.1016/j.cpr.2007.05.006

Holmes, A., Green, S., & Vuilleumier, P. (2005). The involvement of distinct visual channels in rapid attention towards fearful facial expressions. *Cognition & Emotion, 19*, 899-922.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Erlbaum.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching - A review. *Psychological Bulletin*, *136*(5), 849-874.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*, 138-147.

Klinger, E. (1975). Consequences of commitment to and disengagement from incentives. *Psychological Review*, *82*, 1-25.

Koster, E. H. W., Crombez, G., Verschuere, B., Vanvolsem, P., & De Houwer, J. (2007). A time-course analysis of attentional cueing by threatening scenes. *Experimental Psychology*, *54*(2), 161-171.

Koster, E. H. W., Verschuere, B., Crombez, G., & Van Damme, S. (2005). Time-course of attention for threatening pictures in high and low trait anxiety. *Behaviour Research and Therapy*, *43*(8), 1087-1098.

Kruijt, A. W., Parsons, S., & Fox, E. (2019). A Meta-Analysis of Bias at Baseline in RCTs of Attention Bias Modification: No Evidence for Dot-Probe Bias Towards Threat in Clinical Anxiety and PTSD. *Journal of Abnormal Psychology*, *128*(6), 563-573. https://doi.org/10.1037/abn0000406

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL*.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *Das State-Trait-Angstinventar (STAI): Theoretische Grundlagen und Handanweisung (The State-Trait Anxiety Inventory (STAI): Theoretical principles and manual)*. Beltz.

Lipp, O. V., & Derakshan, N. (2005). Attentional bias to pictures of fear-relevant animals in a dot probe task. *Emotion, 5*, 365-369.

Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford handbook of event-related potential components* (pp. 330-360). Oxford University Press.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15-20.

Mair, P., & Wilcox, R. R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*(2), 464-488. https://doi.org/10.3758/s13428-019-01246-w

Mogg, K., & Bradley, B. P. (1998). A cognitive-motivational analysis of anxiety. *Behaviour Research and Therapy, 36*, 809-848.

Mogg, K., Bradley, B. P., Miles, F., & Dixon, R. (2004). Time course of attentional bias for threat scenes: Testing the vigilance-avoidance hypothesis. *Cognition & Emotion, 18*, 689-700.

Mogg, K., Bradley, B. P., & Williams, R. (1995). Attentional bias in anxiety and depression: The role of awareness. *British Journal of Clinical Psychology, 34*, 17-36.

Mogg, K., McNamara, J., Powys, M., Rawlinson, H., Seiffer, A., & Bradley, B. P. (2000). Selective attention to threat: A test of two cognitive models of anxiety. *Cognition & Emotion, 14*, 375-399.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134-140.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97*(2), 316-333.

Onie, S., & Most, S. B. (2022). On the relative sensitivity of spatial and nonspatial measures of attentional bias: Emotion-induced blindness, the dot probe, and gradations in ratings of negative pictures. *Emotion, 22*(8), 1942-1951. https://doi.org/10.1037/emo0000855

10.1037/emo0000855.supp (Supplemental)

Peirce, J. W., Hirst, R. J., & MacAskill, M. R. (2022). *Building Experiments in PsychPy (2nd ed.)*. Sage.

Petrova, K., Wentura, D., & Bermeitinger, C. (2013). What happens during the stimulus onset asynchrony in the dot-probe task? Exploring the role of eye movements in the assessment of attentional biases. *PLoS ONE, 8*, e76335.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160-174.

Puls, S., & Rothermund, K. (2017). Attending to emotional expressions: no evidence for automatic capture in the dot-probe task. *Cognition and Emotion*, 1-14. https://doi.org/10.1080/02699931.2017.1314932

Schmidt, F., Haberkamp, A., & Schmidt, T. (2011). Dos and don'ts in response priming research. *Advances in Cognitive Psychology*, *7*, 120-131.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). North-Holland.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. https://doi.org/10.1037/h0054651

Townsend, J., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory (Vol. 3)* (pp. 200-239). Erlbaum.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Vogt, J., De Houwer, J., & Crombez, G. (2011). Multiple goal management starts with attention: Goal prioritizing affects the allocation of spatial attention to goal-relevant events. *Experimental Psychology, 58*(1), 55-61. https://doi.org/10.1027/1618-3169/a000066

Vogt, J., De Houwer, J., Crombez, G., & Van Damme, S. (2013). Competing for attentional priority: Temporary goals versus threats. *Emotion, 13*(3), 587-598. https://doi.org/10.1037/a0027204

Vogt, J., De Houwer, J., & Moors, A. (2011). Unintended Allocation of Spatial Attention to Goal-Relevant but Not to Goal-Related Events. *Social Psychology*, *42*(1), 48-55. https://doi.org/10.1027/1864-9335/a000042

Vogt, J., De Houwer, J., Moors, A., Van Damme, S., & Crombez, G. (2010). The automatic orienting of attention to goal-relevant stimuli. *Acta Psychologica*, *134*(1), 61-69. https://doi.org/10.1016/j.actpsy.2009.12.006

Vogt, J., Dodd, H. F., Parker, A., Duffield, F., & Sakaki, M. (2022). Now you see it, now you don't: Relevance of threat enhances social anxiety-linked attentional bias to angry faces, but relevance of neutral information attenuates it. *PLoS ONE*, *17*(7), e0271752. https://doi.org/10.1371/journal.pone.0271752

Vogt, J., Koster, E. H. W., & De Houwer, J. (2017). Safety first: Instrumentality for reaching safety determines attention allocation under threat. *Emotion, 17*(3), 528-537. https://doi.org/10.1037/emo0000251

Vogt, J., Lozo, L., Koster, E. H. W., & De Houwer, J. (2011). On the role of goal relevance in emotional attention: Disgust evokes early attention to cleanliness. *Cognition & Emotion, 25*(3), 466-477. https://doi.org/10.1080/02699931.2010.532613

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (pp. 95-116). Guilford.

Wentura, D., Messeh, L., & Wirth, B. E. (2024). Attentional bias towards happy faces in the dot-probe paradigm: it depends on which task is used. *Cognition & Emotion, 38*(2), 217-231. https://doi.org/10.1080/02699931.2023.2283014

Wentura, D., & Wirth, B. E. (2023, August 16). Replication Vogt et al. (2013). *Retrieved from osf.io/z52hc*.

Wentura, D., Wirth, B. E., & Pospeschill, M. (2023). *Multivariate Datenanalyse mit R (2. überarbeitete Aufl.)*. Springer. https://doi.org/10.1007/978-3-531-93435-8

Wilcox, R. R. (2013). *Introduction to robust estimation and hypothesis testing (3rd ed.)*. Academic Press.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1988). *Cognitive psychology and emotional disorders*. Wiley.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1997). Cognitive psychology and emotional disorders (2nd ed.).

Wirth, B. E., & Wentura, D. (2018a). Attentional bias to threat in the general population is contingent on target competition, not on attentional control settings. *Quarterly Journal of Experimental Psychology*, *71*(4), 975-988. https://doi.org/10.1080/17470218.2017.1307864

Wirth, B. E., & Wentura, D. (2018b). Furious snarling: Teeth-exposure and anxiety-related attentional bias towards angry faces. *PLoS ONE*, *13*(11), 14, Article e0207695. https://doi.org/10.1371/journal.pone.0207695

Wirth, B. E., & Wentura, D. (2019). Attentional bias towards angry faces is moderated by the activation of a social processing mode in the general population. *Cognition & Emotion, 33*(7), 1317-1329. https://doi.org/10.1080/02699931.2018.1561423

Wirth, B. E., & Wentura, D. (2020). It occurs after all: Attentional bias towards happy faces in the dot-probe task. *Attention Perception & Psychophysics, 82*(5), 2463-2481. https://doi.org/10.3758/s13414-020-02017-y

Wirth, B. E., & Wentura, D. (2023). Social processing modulates the initial allocation of attention towards angry faces: Evidence from the N2pc component. *Social Cognitive and Affective Neuroscience, 18*(1). https://doi.org/10.1093/scan/nsad070

Yiend, J., & Mathews, A. (2001). Anxiety and attention to threatening pictures. *The Quarterly Journal of Experimental Psychology*, *54A*, 665-681.

Appendix A

Stimuli used in the main phase

Table A1

Stimuli used in the main phase (IAPS# in parentheses)

Type	Set A	Set B
Neutral/Goal[a]	Mug (7009)	Bowl (7006)
	Book (7090)	Mushroom (5510)
Threat-related	Attacking snake (1120)	Snarling dog (1300)
	Knife attack (6350)	Gun attack (6560)
Neutral (only goal task)	Rolling pin (7000)	Fork (7080)

[a] For each participant, one picture was the goal stimulus (counter-balanced).

Appendix B

Deviations from the preregistration

In the preregistration, we described a single experiment with a 2 (*task*: location vs. discrimination) × 3 (*trial type:* goal picture/neutral picture vs. goal picture/threatening picture vs. threatening picture/neutral picture) × 2 (*cue validity*: invalid vs. valid cue) design with task as a between-participants factor and trial type and validity as within-participants factors. Why did we split up the report into two experiments?

From the start on the essential question of our research was whether the effect of attentional priority for temporary goals survives in a non-confounded procedure (i.e., with the discrimination task). Therefore, the core of the project was an experiment using a discrimination task and to see whether a substantial effect of attentional priority will be found with the improved design. We did this successfully with our Experiment 2.

Of course, it was necessary to complement this experiment with a replication of the original experiment with the location task for two reasons. Most important, in order to be unambiguously interpretable, a possible null result on the discrimination task would have had to be contrasted with a clear replication on the location task. Therefore, Experiment 1 was needed to show the replicability of the original result to provide the background for our main experiment (i.e., Experiment 2). On a minor note, in case of significant effects in both experiments, an overall analysis reveals potential significant quantitative differences.

However, in the latter case in particular, a report as a single experiment set the wrong emphasis: now, the quantitative differences in cueing effects of the two tasks are in the foreground and the most important result – that is, the essential hypothesis is still supported if the confound has been removed by using the adequate task – is deemphasized.

Is this deviation from the preregistration a "questionable practice"? No, not in this case. We preregistered (emphasis added): "The data will first be analyzed with a 2 × 4 × 2 mixed-design MANOVA. … If this MANOVA reveals any significant interaction effects, follow-up t-tests will be calculated to clarify these interactions. Most importantly (*and independently of the outcome of the MANOVA*), we will conduct t-tests in order to assess the cue validity effect for dot-probe trials where a goal picture and (a) a neutral and (b) a threatening picture are presented as cues (*separately within each group of participants*)."

Thus, it was a priori planned to report results separately for the two partial experiments. Thus, the only deviance from the preregistration is that we additionally report separate MANOVAs for the two experiments *before* the report of an overall MANOVA that justified the separate MANOVAs as follow-up analyses. Table B1 lists other minor deviations from the pre-registration.

Table B1

Other minor deviations from the pre-registration

Preregistration	**Deviation**	**Reason**
"The data will first be analyzed with a 2 × **4** × 2 mixed-design MANOVA."	We analyzed (in the cross-experiment analyses of this article) with a 2 × **3** × 2 mixed-design	In the preregistration, we accidentally wrote about a 2 × **4** × 2 analysis, that is, we included the filler trial type (threat/threat; see *Procedure* of Experiment 1) as a fourth condition. This is not meaningful because in the filler trials the conditions "valid" and "invalid" are defined only at the level of the experimental program and are equally occupied by two threat-related pictures. Thus, the "cueing effect" for fillers should *a priori* not be deviant from zero (which is the case here; $t[34] = 1.55$, $p = .130$, $d_Z = 0.26$ and $t[29] = -1.39$, $p = .174$, $d_Z = 0.25$, for Experiment 1 and 2, respectively).
"In order to assess the significance of specific effects, the Pillai's trace criterion will be used."	We did not report Pillai's trace criterion.	It was a mistake on our part to mention Pillai's trace criterion here, as it implies that it would have made a difference which multivariate criterion (i.e., Pillai, Wilks, Hotelling-Lawley, Roy) had been chosen. That is not true in the given case. (The multivariate criteria differ only if the number of conditions of the within-participants factor and the number of conditions of the between-participants factor are both larger than two; see, e.g., Wentura et al, 2023.) Note: We report η_p^2 as an effect size; η_p^2 is identical with Pillai in the given case.

Appendix C

Inverse Efficiency Scores (exploratory)

We added an analysis of inverse efficiency scores (IES; Townsend & Ashby, 1978). IES are defined by mean RTs divided by the proportion of correct responses. As processes underlying effects in reaction time tasks can be reflected in RTs or error rates, combining them can yield a clearer picture onto the underlying process. Note, Bruyer and Brysbaert (2011) did not unconditionally recommend this measure. However, in (a) the case of rather low error rates (< 10%) and (b) a clear positive correlation of mean RTs and error rates across the conditions (i.e., no evidence of a speed-accuracy tradeoff) it might be useful. Both characteristics are given here (see Table 1). A further cautionary remark given by Bruyer and Brysbaert is the fact that IES increases the variability of the data. The procedure can especially yield outlier values: if, e.g., in a specific condition the mean RT of three participants is 500 ms, but the accuracy rates are .95, .75, and .50, the IES scores are 526, 667, and 1000. Table C1 shows the IES and cueing scores based on IES. In fact, boxplots of the cueing scores indicate the outlier problem mentioned above; in Table C1, this can be seen by the exaggerated standard errors. Therefore, we reported trimmed means as well (see Table B1, right column) and analyzed the cueing scores with robust tests. We tested individual cueing scores with robust one-sample t-tests (function yuen.t.test from the R package PairedData; Champely 2018; see Wilcox, 2013, with regard to robust testing) with a trimming of $\gamma = .20$. We compared cueing scores between the task samples with robust t-tests for independent samples (function yuen from the R package WRS2; Mair & Wilcox, 2020). Due to the peculiarities of the IES mentioned above, we do not consider the following analyses to be a substitute for the RT analyses reported in the main text. They should be taken as a further piece of information to evaluate the comparison of the location and discrimination task.

Table C1

Mean Inverse Efficiency Scores (IES) and Cueing Scores as a Function of Cue Validity, Cue Type, and Task

Task	Cue validity		Cueing-Score	
Trial type	Valid	Invalid	Mean	Trimmed Mean
Location Task				
Goal vs. Neutral	457	518	61 [12]	54 [8]
Goal vs. Threat	460	529	69 [9]	62 [6]
Threat vs. Neutral	484	484	-1 [7]	-5 [7]
Discrimination Task				
Goal vs. Neutral	635	677	42 [32]	24 [14]
Goal vs. Threat	616	683	68 [19]	53 [17]
Threat vs. Neutral	637	639	2 [11]	-6 [11]

Note. Cueing score = $IES_{nvalid} – IES_{valid}$; SEs are given in brackets. Discrepancies between mean IES and mean cueing scores are due to rounding. Trimmed means are based on a trimming of γ = .20; the SE of the trimmed mean was calculated by dividing the trimmed mean by the t-value of a robust t-test. “Valid” refers to trials in which the target replaced the cue category first mentioned under trial type; “invalid” refers to trials in which the target replaced the cue category mentioned second under trial type.

For goal/neutral trials, the pattern corresponds closely to what we already knew from the RT analyses: numerically, the cueing score drops to less than half from location task (M_{tr} = 54) to discrimination task (M_{tr} = 24). The mean cueing score for the location sample is significant, $t(20) = 6.64$, $p < .001$, $d_Z = 1.20$, whereas the mean cueing score for the discrimination sample even missed the criterion of significance, $t(17) = 1.75$, $p = .098$, d_Z = 0.34. (However, from a replication perspective, a one-tailed test might be seen as appropriate here, hence p = .049.) The difference in mean cueing scores is associated with $t(28.28) = 1.89$, $p = .069$, $d_S = 0.37$. The failure to meet the conventional criterion of significance should be attributed to the still large standard error of the cueing score for the discrimination task.

The picture is somewhat different for goal/threat trials. First, the (trimmed) mean IES cueing scores are rather close ($M_{tr} = 62$ vs. $M_{tr} = 53$); both mean cueing scores were significant, $t(20) = 9.53$, $p < .001$, $d_Z = 1.72$ for the location sample and $t(17) = 3.21$, $p = .005$, $d_Z = 0.63$, for the discrimination sample. Second, the difference in mean cueing scores is now clearly not significant, $|t| < 1$.

For the sake of completeness, for threat/neutral trials neither the difference in mean cueing scores nor the task-specific cueing scores were significant, all $|t|$s < 1.