Searching for a face in the crowd: Pitfalls and unexplored possibilities

D. Vaughn Becker[1] · Hansol Rheem[1]

Abstract
Finding a face in a crowd is a real-world analog to visual search, but extending the visual search method to such complex social stimuli is rife with potential pitfalls. We need look no further than the well-cited notion that angry faces "pop out" of crowds to find evidence that stimulus confounds can lead to incorrect inferences. Indeed, long before the recent replication crisis in social psychology, stimulus confounds led to repeated demonstrations of spurious effects that were misattributed to adaptive cognitive design. We will first discuss how researchers refuted these errors with systematic "face in the crowd" experiments. We will then contend that these more careful studies revealed something that may actually be adaptive, but at the level of the signal: Happy facial expressions seem designed to be detected efficiently. We will close by suggesting that participant-level manipulations can be leveraged to reveal strategic shifts in performance in the visual search for complex stimuli such as faces. Because stimulus-level effects are held constant across such manipulations, the technique affords strong inferences about the psychological underpinnings of searching for a face in the crowd.

Keywords Visual search · Anger superiority effect · Emotional expression · Face processing

Any cognitive scientist who has lost a set of keys in a cluttered room realizes that there is more to the act of visual search than can be revealed with simple stimuli like colored letters. We can of course recognize the tremendous progress made by Treisman's visual search methodology using such simple stimuli. The approach has been exemplary of the strong inference that the best cognitive science affords. Systematic manipulations of color, form, and other stimulus features have teased apart the role of attentional strategies, unveiling whole new perspectives on how we find objects in a complex world. As with many of the methods of cognitive psychology, the simplicity and controllability of the stimuli affords systematic revelations about process—in this case, of early visual attention.

Given that our cognitive capacity for visual search almost certainly reflects a system that was selected to handle messier tasks, it was inevitable that more ecologically valid stimuli would be employed. Visual search for faces was a natural extension of the technique, and facial expression of anger proved an intuitively compelling signal to examine. Many fell under the spell of a beautiful idea—that our attentional system had evolved to detect threats such as angry faces as readily as it does color contrasts (e.g., Hansen & Hansen, 1988; Öhman, Lundqvist, & Esteves, 2001). Study after study purported to show that angry faces were found quickly and efficiently, regardless of the size of the crowd. Indeed, some studies even purported to show that angry faces "popped out" like a red singleton in a field of green distractors. On the other hand, a growing body of research contests these findings, implicating low-level (nonemotional) facial features as the true drivers of efficient searches for angry faces. Despite this mounting accumulation of ugly empirical facts, it is still not clear that we have slain the seductively beautiful notion that angry faces "pop out" of crowds.

When the first author began investigating the issue, early demonstrations of what became known as the anger superiority effect (ASE) were already making waves in social and evolutionary circles. This is not surprising: Putting angry faces into the visual search task seems an ideal intersection of cognitive methods and social content for any cognitive psychologist interested in evolutionary ideas. Replications of the ASE using simple schematic angry faces accumulated. Unfortunately, replicating the ASE proved highly contingent on the stimuli used, and many failures to replicate also arose.

✉ D. Vaughn Becker
    vaughn.becker@asu.edu

[1]   Arizona State University, Tempe, AR, USA

The reason for these discrepant findings is that the visual search methodology has certain "loopholes" that can be exploited in the face of complex stimuli. For example, it allows participants who are miserly with their cognitive resources (e.g., college freshman working for Intro Psych credit) to discover ways to perform the task that do not involve deep levels of stimulus processing. Indeed, if detecting a simple feature that is part of the more complex stimulus allows participants to do the task quickly, they may not even process the emotion of the face. The task can be quite boring when one is looking at similar displays over hundreds of trials, and this induces all but the most compliant participants to search for low-level stimulus features to comply with what they perceive to be the experiment's requirements (e.g. "detect as quickly and as rapidly as you can"). So, although exact replications of the most cited ASE manuscripts did indeed produce similar results, this seems to be because they also replicated confounds that drove those findings.

Here, we will review some of the field's forays into the visual search for expressive faces. In the process, we will develop suggestions for better applications of the visual search to complex stimuli such as faces, with lessons that may help a new generation of researchers avoid some all-too-common pitfalls. We will also explore how the visual search task might afford inferences about the signal form of another expression: happiness. Finally, we will suggest promising ways that visual search for threats (and other complex stimuli) might be leveraged to reveal individual differences in threat detection, provided we avoid the pitfalls and confounds that have befallen past explorations.

Treisman's visual search methodology

The visual search task (Treisman & Gelade, 1980) requires participants to search through arrays of objects for a single target item. It can reveal differences in the efficiency with which particular kinds of targets can be found. For many kinds of stimuli, we must focus attention on a specific location in order for the features of that object to be bound together. The size of the array—and thus the number of distractors—is a key manipulation because it allows us to estimate the rate at which attention can bind an object's features together, evaluate them, and detect or reject these objects relative to the task goals. The search rate should thus vary as a function of the number of distracters. The slopes of these functions indicate how much an additional object to the array affects overall search time. The steeper the slope, the longer the processing time for each additional distractor.

There are cases in which a target can be detected in the same amount of time regardless of the number of distractors. An "X" in a field of "O"s, for example, might be found just as rapidly in a field of 20 distractors relative to a field of only

four. The form "pops out" at us. Such results suggest that mechanisms can guide attention to the target's location without spatial attention having to being deployed: a *parallel search* (Wolfe, 1992).

Of course, true parallel searches are not as common as searches that do require the serial deployment of attention. Figure 1 contrasts the pattern that indicates a parallel search process versus two varieties or serial search processes. If stimuli do not "pop out" because of parallel processes, participants must use the serial deployment of attention to various locations. They will search these arrays, one item at a time, and terminate the search when the target is found. The slope of such a search will thus be half that of the exhaustive search. This has traditionally been called a *serial self-terminating* search. Some simple features—such as color, motion, and shape—do induce efficient parallel searches (e.g., see Table 13.1 of Wolfe, 2018). The top panel of Fig. 2 shows a search for an "X" in an array of "O"s across two array sizes, and it should be apparent that this search will be very efficient. But now consider looking for the same target—a gray "X"— in an array of Black "X"s and gray "O"s. Treisman and colleagues showed that conjunctions like this, of two of simple features, which on their own pop out, suffer with increasing numbers distractors, provided those distractors each share one of the two features that make up the conjunction. Most feature conjunctions and other more complex targets require serial self-terminating searches. They have target detection slopes that differ significantly from zero, and the value of these slopes is a function not only of the target attributes but also of how efficiently we can process and reject the distractors. We may nevertheless encounter complex stimuli that might have simple features that drive greater efficiency. When we consider complex stimuli like those in the lower panel of Fig. 2, it is not difficult to imagine that all of the complexity may be trumped by the stimulus outlines, yielding search patterns that look a lot like the "X"s and "O"s of the upper panel. If we saw this threatening stimulus give rise to flat slopes (i.e., "pop out"), we should be cautious about attributing this to the threat and not the shape. Although the parallel/serial distinction may now be considered overly simple (e.g., Wolfe, 1998; Wolfe, & Horowitz, 2004), it serves well to illustrate the potential pitfalls that can arise in the visual search for complex images.

Searching for a face in the crowd

When we consider faces, many would maintain that they are conjunctions of features. Consequently, a visual search for a face with a particular identity or attribute should be laborious and happen in serial. However, given that our visual system has coevolved with facial signaling systems, it is possible that special feature detectors keyed to important facial signals might exist. Moreover, many have argued that faces are
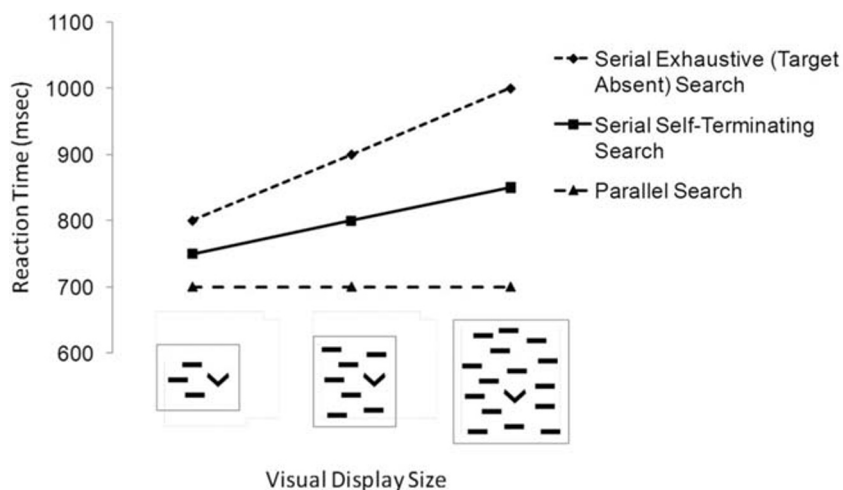
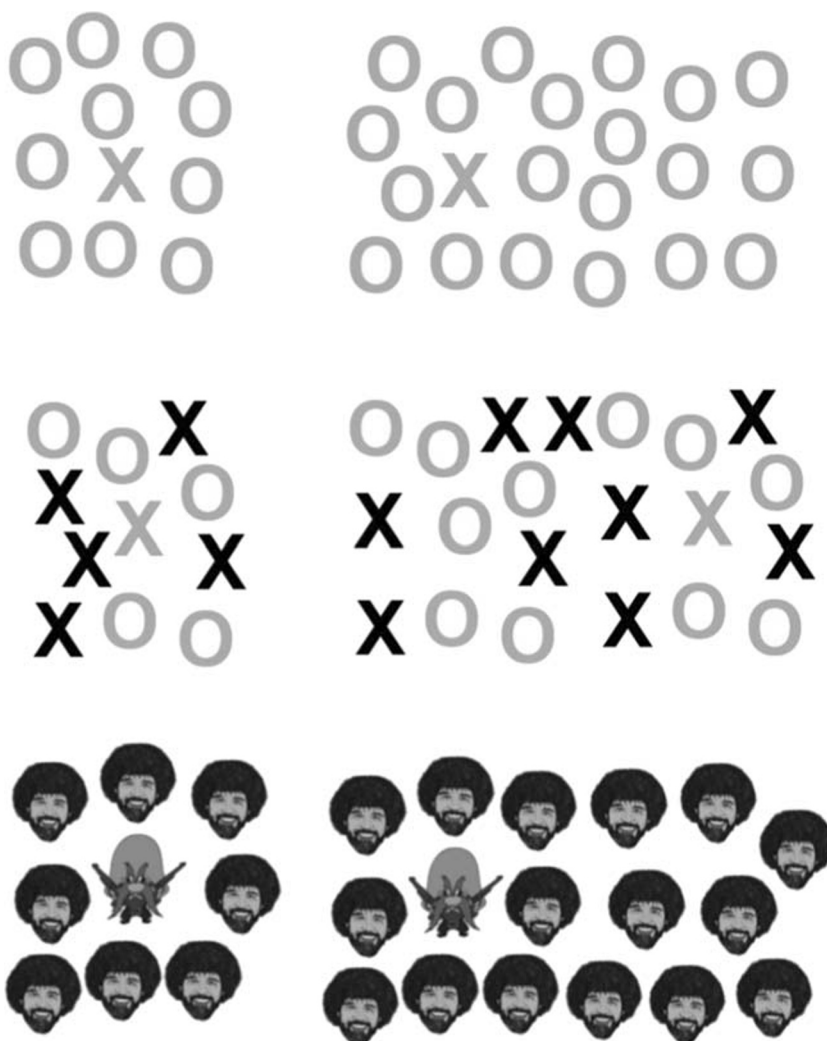Fig. 1 Varying the array size n visual search allows researchers to discriminate between serial and parallel searches

Fig. 2 What kinds of stimuli yield which kinds of searches? The top panel shows that the form of the "X" among distractors of a different form can be detected in parallel. The middle panel in contrast shows a search for the same target, but the distractors now share either the target's form or color, which requires the serial deployment of attention to detect.

Which of these patterns will arise for the even more complex angry and happy stimuli in the lower panel? Will attention need to be deployed in a serial self-terminating fashion, or will the salience of the shapes produce a pattern more akin to a parallel search (even though the happiness and anger are incidental to this task)?

perceived holistically (Taubert, Apthorp, Aagten-Murphy, & Alais, 2011)—gestalts that are more than the sum of their parts. From such perspectives, certain configural aspects of faces might be detected efficiently, and even give rise to "pop-out" effects. What kind of facial attributes are good candidates to reveal such efficient searches?

Hansen and Hansen (1988) seem to be the first to hypothesize that nature selected cognitive mechanisms to facilitate the early detection of angry faces. Their adaptationist logic is intuitively satisfying to many. Emotional expressions such as anger convey messages with important—potentially dire— consequences for the receiver's fitness (Fridlund, 1994). This provided an impetus for selection to favor mechanisms that hasten their detection. Nature might even have crafted feature detectors dedicated to this task, perhaps a subset of Selfridge's pandemonium demons, each monitoring a small share of the visual scene, and ready to shriek if an angry face is seen (Selfridge, 1959).

Many now postulate mental mechanisms that automatically scan the environment for threats and which then reallocate cognitive resources to better process and respond to these dangers (Öhman & Mineka, 2001). The facial signal of anger communicates displeasure and goal frustration on the part of the displayer, and often warns of impending aggression and violence (Scherer & Wallbott, 1994). It is thus plausible that the angry face is in the class of attentionally privileged objects, and that it might rapidly draw attention to itself even when hidden in large crowds. There is plenty of ancillary evidence that seems to support such reasoning. For example, angry expressions are detected with considerable levels of accuracy across cultures (Ekman & Friesen, 1971[1]). This universality of the signal form of anger may suggest that nature has had the time to select for specialized feature detectors. Of course, it is equally important to recognize that natural selection does not optimize, but merely satisfies. It is a bricoleur, using whatever materials are available to satisfy the functional goal. In other words, just because something would be adaptive does not mean an adaptation has arisen, especially when other factors allow a simpler solution. Still, many other pieces of evidence do suggest efficiencies in anger detection. For example, studies using classical conditioning have shown that angry faces can be subliminally perceived (Dimberg & Öhman, 1996; Öhman & Dimberg, 1978). Of course, the fact that participants gaze directly at the faces in such studies precludes strong claims about an angry face automatically drawing attention to its location *prior* to a controlled search. Something like the visual search method is needed for such inferences (Fig. 3).

Hansen and Hansen (1988) attempted to provide this evidence: They used four and nine face crowds in a visual search task to show that anger "popped out" of displays. Given that

they did this in the early days of personal computer monitors, they can be forgiven for using thresholded bitmaps[2] of the classic Ekman and Friesen (1976) faces. Unfortunately, this introduced the possibility of low-level confounds. Such potential confounds could be even more salient because the crowds they used were utterly homogenous, potentially confounding target detection with distracter rejection. The same specific angry target was used repeatedly as well, opening up the possibility that participants could learn to efficiently search for these targets using some artifact of the image thresholding process. Although such methodological concerns were known, they were underappreciated by early audiences outside of the cognitive field. The intuitive appeal of the finding that angry faces "popped out" were embraced by many and citations accumulated rapidly.

Despite the enthusiasm, some cognitive psychologists were concerned that spurious factors may be at work. Purcell, Stewart, and Skov (1996) mounted careful empirical explorations of the anger superiority effect (ASE). Using the original images, they were able to show that the bitmap thresholding procedure had indeed introduced a low-level feature—a black blotch—into the base of the angry face. This allowed for very efficient searches that had nothing to do with angry face detection. Purcell and colleagues not only replicated the confounded search, they then removed the black blotch and showed that the ASE went away. Then they added back in a white blotch and demonstrated the pop-out effect again. This proved a powerful demonstration that it was not anger that popped out, but a blotch.

This failure to replicate occurred at a time when psychological science was less vigilant about replicability issues, and so it may not be surprising to learn that it took several years and rounds of less than pleasant reviews for Purcell and colleagues to get this published (Purcell, personal communication). And in the interim, the idea of the ASE had continued to gain steam. Other demonstrations soon arrived using *schematic* stimuli (e.g., Fox et al., 2000; Hahn & Gronlund, 2007). These stimuli were more controlled, but in their simplicity were also potentially confounded by low-level features. Some of these schematic faces (e.g., Eastwood, Smilek, & Merikle, 2001, 2003) were not even clearly angry—they were frowny faces, looking more sad than anything—and it is harder to make the case that it would be adaptive to detect sad faces. Many were clearly angry, however, and those developed by Öhman et al. (2001) fell into this category. Across five experiments, they produced what they took as evidence of the ASE, and their manuscript became the most

---

[1] Although, note that the accuracy is nowhere near that with which happy facial expressions are identified.

[2] For younger readers: Early computer monitors had very low resolution and were incapable of displaying photographic images with any great detail. One either used photographs and tachistoscopes, or one reduced the image complexity—turning every 10×10 patch of grayscale pixels into a black or white super-pixel appropriate to monitor resolution—while trying not to do too much aesthetic violence to the image in the process.

Fig. 3 In early demonstrations that anger "popped out," low monitor resolution required "thresholding" of photographs that introduced low-level vision confounds. Over hundreds of nearly identical trials, participants could readily learn to search for the dark blotch at the bae of the angry face to complete the task. Anger did not "pop out"—darkness did

widely cited replication of the ASE. Their stimuli have been widely used since.

Many aspects of their design conformed to typical visual search designs. They used different crowd sizes to assess the search rates. Some of the experiments used neutral faces for the crowds, which held constant the search rates though the distractor faces. But there were also some troubling symptoms of potential problems with Öhman et al.' (2001) stimuli. For example, in their Study 4, they flipped the faces upside down—a manipulation that normally disrupts facial processing—and still observed the ASE. An adaptation to detect anger even when the attacker is upside down seems unlikely to have posed a sufficient threat to our ancestors to exert the requisite selective pressure to create a dedicated feature detector. These effects seem instead to reflect more general issues with these stimuli, and are similar to those in the Hansen and Hansen (1988) studies: Targets had potential confounds, and homogenous crowds often confounded distracter rejection with target detection. Purcell and Stewart (2010) were again able to find and empirically demonstrate faults with the angry schematic faces: the angry eyebrows in Öhman's schematic faces proved a confound. Specifically, the configuration of elements radiating out from the center of the angry face (but not the happy or neutral faces) were more rapidly detected even when they were not understood to be faces. If the features themselves pop out without being in a face-like context, it undermines any inferences that facial threat is driving the result, its more parsimonious to blame the feature. Of course, just because some participants might

neglect instructions to look for emotion and rely instead on simpler features, this does not mean that all participants do. Moreover, these schematic faces may well reveal attentional effects in other paradigms, such as the dot probe that examines how attention is held by anger (e.g., Fox, Russo, & Dutton, 2002). Nevertheless, the case for the preattentional ASE should not rest on such evidence. More realistic angry targets embedded in heterogeneous crowds are necessary, and by the late 1990s, monitor resolutions made this possible.

Many authors began using realistic faces. Some of these (e.g., Williams et al., 2005) provided some support for the ASE. Others (e.g., Horstmann & Bauland, 2006) produced what looked like ASEs, but when exhaustive search rates were taken into account, revealed instead a slower search through angry crowds that only made it look like happy faces were detected more slowly than angry faces were (see also Horstmann, Scharlau, & Ansorge, 2006). Saying that we detect angry faces preattentionally is one thing, but a slow search through angry crowds—which slows the detection of the happy target embedded in this crowd—is really making a claim about a serial search process having a difficult time disengaging from angry faces. That is not a parallel search process that functions preattentionally, but rather a drag on attention once it has been serially allocated to a location, a difference in processing and discriminating angry faces from the sought-for target.

Even more problematic was that many investigations using photographs of faces produced something quite different: Efficient searches for *happy* faces (e.g., again, Williams

et al., 2005). To be clear, few of these showed that happy faces popped out of crowds, but they did show a search asymmetry favoring happy faces relative to angry faces, and this obtained even when the exact same heterogeneous crowds of neutral faces were used. The work of Horstmann (2007, 2009) and colleagues (Horstmann, Becker, Bergmann, & Burghaus, 2010) is exemplary in teasing apart the effects of distractor rejection and target detection.

Like these other researchers, for years we also sought and failed to obtain evidence for the ASE using realistic faces. Like many of these other labs, the Happy Superiority Effect (HSE) kept appearing as the more reliable effect. As this picture became clearer, we moved to publish a series of studies that added what we hoped to be decisive evidence against the ASE (Becker, Anderson, Mortensen, Neufeld, & Neel, 2011).

We first used six Ekman and Friesen (1976) stimuli in two, four, and six face crowds to show the HSE consistently triumphed over the ASE. But there was a problem, another potential confound: All but one of the happy faces showed the exposed white teeth of a toothy grin, which may itself come to be learned as a stimulus feature that can drive efficient search as the task wears on. We therefore removed the lower half of the faces in another study, which should have eliminated this source of the effect, but we continued to see the HSE.

Another study used computer-generated stimuli with closed mouth expressions. These had the control of schematic stimuli, but with enough variations in facial shape and lighting that the crowds and targets had a realistic level of heterogeneity, and a consistent number and degree of changes distinguishing happy and angry from neutral distractors. Again, the HSE was evident.

Perhaps the most important piece of evidence Becker et al. (2011) offered against the ASE came from the use of a multiple target visual search task (based on Thornton & Gilden, 2007). If one is looking for an angry face in a small crowd, and more than one stimulus is angry, and if parallel detection processes exist that can facilitate this search, they should speed up the decision to say that an angry face is present relative to arrays in which only a single angry target is present. This is called redundancy gain, and in fact seems to be the best way to detect whether facial displays of anger can be detected without the serial application of visual attention. Consistent with our other work, we again failed to see angry faces efficiently detected relative to happy faces. Nor was there any evidence of angry faces yielding redundancy gains. (There was, however, some evidence for redundancy gain for happy faces—the reaction times showed this, but given a concomitant increase in errors, we cannot rule out speed–accuracy trade-offs.)

Another finding of the Becker et al. (2011) redundancy gain work reveals an underappreciated possibility for visual search studies using complex stimuli like faces. One set of participants detected angry and happy targets in arrays of neutral faces, while another set detected the same targets in arrays of fearful faces. Global performance slowed down when the distractor images were fearful versus neutral. Importantly, these slowdowns occurred for trials with redundant displays, in which all of the faces were angry (or happy). The neutral or fearful distractors were only being seen on other trials when fewer targets were present or targets were absent entirely. In other words, the exact same displays—four example, four angry faces—yielded significantly longer RTs (and accuracy does not trade off to account for these effects) in one condition compared with the other. Why would performance be affected by trial types (fearful and neutral crowds without any targets) that are happening at other times in the experiment? Consistent with work in semantic (and other) priming research, this seems to be a case in which different kinds of distractors are raising the level of evidence that participants need to identify a target. Consider the lexical decision tasks, where words must be discriminated from nonwords. It has been widely shown that performance on one trial (e.g., identifying the word "DOCTOR") can speed performance on a subsequent trial to a related word (e.g., "NURSE"). Importantly, however, this effect is modulated by the kinds of nonwords that are used. If "NURSE" must be discriminated from unpronounceable nonwords like "XLFRT," the task is easier relative to nonwords have to be pronounced ("GERT"; Stone & Van Orden, 1993; see also Becker, Goldinger, & Stone, 2006). This later variation will slow performance and enhance the magnitude of priming effects. We suspect something similar in explaining our redundancy gain results: The need to discriminate angry from fearful faces on some trials slows performance even when no distractors are present. Note again that this entails an effect that is not based on the stimulus display itself (for they are constant) but rather on some criterion shift internal to the participant based on other trials or parts of the experimental session. This is a promising methodology for future investigations of the visual search for faces, and we will discuss it again in the final section.

It is worth acknowledging that none of this completely rules out the possibility of the ASE. For example, static images are a far cry from the dynamic facial displays of rage and other emotions. Dynamic enraged faces may well "pop out" of crowds of gently smiling distractors. But does that say something about the perceiver, or the signal itself? It should also be noted that some individual differences might hasten the detection of anger. For example, Byrne and Eysenck (1995) showed that anxious participants more efficiently detected angry faces (few who cite this study also note that they found a main effect that favored happy face detection).

Despite these failures to replicate, there continues to be a steady stream of new researchers who wish to further explore the detection of expressive faces in visual search. We here note

several recommendations for such efforts (consistent with those offered by Becker et al., 2011, and Frischen, Eastwood, & Smilek, 2008):

- First, if one wants to make inferences about processing asymmetries in search rates, set size must be varied within participants in order for slopes to be calculated. The multiple target search methodology may be an even more effective design.

- Second, these detection slopes need to be assessed relative to the target-absent searches. As noted above, analyses in terms of such search ratios are common practice in the visual search literature (see Wolfe, 1992) and can readily disambiguate whether targets are efficiently detected controlling for the rates of search through the distractors.

- Third, the distractor crowds should be held constant across the different types of targets. If angry faces appear to be more efficiently detected relative to happy (or other expressive) faces, but the angry faces occur within happy crowds while the happy faces appear within angry crowds, target detection effects are confounded with different types of crowds. Frischen et al. (2008) further advocate that nonexpressive faces be used as distractors for all target-detection trials. We would caution that even here, discriminability of the different targets from the common distractors could confound the results (see below).

- Fourth, studies must be designed so that participants are consciously searching for particular kinds of expressions and rule out the possibility that some nonemotional feature common to the target types can be learned and come to drive an efficient search.

- Fifth, it is also prudent to recommend that many different targets and distractors are used, and that when possible, designs are replicated with different stimulus sets that still satisfy whatever stimulus considerations are under investigation.

- Sixth, symmetrically spaced crowds can sometimes allow participants to use the space between items to detect targets. Jittering the positions of the items in the crowds ensures that textural gestalts cannot be exploited to complete the task, because again, this has nothing to do with detecting threats.

We hope these guidelines will help those continuing to look for face in the crowd effects. We now turn to the positive finding that happy faces are more efficiently detected.

The happy superiority effect

When we look at the more efficient detection of happy faces—the HSE—we should be just as cautious as we are with angry faces. Even more so, because while the adaptiveness of

detecting someone with aggressive intentions is obvious, the adaptiveness of detecting someone that wants to be your friend seems valuable, but it is something that can be accomplished at a more leisurely pace. Indeed, there are many studies that reveal happy-face advantages at later stages of processing such as recognition memory (some are reviewed in Becker & Srinivasan, 2014). In the early stages of information processing, however, a search asymmetry favoring happy faces appears to be increasingly reliable. Of course, if the happy expressions have exposed teeth, it is tempting to dismiss such demonstrations, because like salient angry eyebrows, the contrast of the white teeth with the rest of the face provides a visual feature that attention can get a grip on, even if the emotional significance is not recognized. This would be similar to spurious effects elicited by the black blotch in the angry face photo used in Hansen and Hansen's (1988) original study of the ASE. However, there is a big difference in the reason these features exist in the experimental stimuli. The schematic eyebrows are a function of experimenter choice about how to operationalize emotional meaning without sacrificing stimulus control. Toothy grins, in contrast, are *not* an experimenter choice—they are a fundamental aspect of how humans express happiness and prosocial intentions.

This changes the kinds of inferences that we can make, because it raises the question of *why* we express happiness with our teeth. When most animals display their teeth, they are not being prosocial—it is a threat. So how did we come to use the teeth to signal something so different, a prosocial intention? Unlike most animals, we walk upright and show our face to the world, and this has allowed it to become a much richer conveyor of signals. The form of these signals would have coevolved with the feature detectors that were already present in the visual system. The form of facial signals would be constrained by receiver capacities, but it would also take advantage of them. We have speculated (Becker & Srinivasan, 2014) that the social nature of our hominid ancestors may have provided the preconditions for the toothy signal of threat to become one of positive affordance, precisely because contrast detectors allow it to be more efficiently detected. In other words, we can still use teeth in an enraged facial expression to better signal threat, but we now also use the contrast of the teeth with a different facial framing to more efficiently signal that we are not a threat.

The reason for the happy face's salience in a crowd may be a by-product of two other adaptive benefits toothy grin confers: discriminability at a distance and discriminability at brief exposures. Hager and Ekman (1979) first showed that happiness is better detected at distances beyond 50 meters, where other expressions cannot be resolved. In fact, the toothy grin can convey its message from outside the distance of an Olympic javelin throw (>90 meters). In other words, the toothy grin might now signal prosociality (rather than only rage) precisely because it was discriminable enough to defuse

intertribal conflicts outside of projectile weapon distances in the ancestral past. This would provide an advantage for which nature might have selected, but there is another situation that would also facilitate the adoption of this signal form: signaling prosociality and submission *rapidly*. Facial happiness can be discriminated from other expressions at as little as 27 ms of exposure duration, a level at which the expressions of fear, anger, and neutrality are effectively erased by a poststimulus mask (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007). Thus, at distances within the range of thrown weapons, this is a good way to rapidly defuse social threats (including angry faces). There is a precedent in the chimpanzee facial expression of submission, which also uses exposed teeth in an open-mouthed expression. The flash of teeth is a powerful contrast that renders any signal more discriminable, and in our closest ancestors, the chimpanzees, it has been appropriate to signal both rage and submission. The other aspects of the signal discriminate between these two signals, and the human signals have recruited even more facial musculature to further discriminate them. Let us be clear about our argument: It is not that feature detectors evolved to detect toothy grins, but that behavior of making toothy grins was selected because it appealed to preexisting feature detectors and better preempted conflict.

Thinking about the coevolution of facial signals and their detectors may explain why static displays of anger are not efficiently detected in visual search. Selection for signal forms would have been especially effective if the communication was beneficial for both the signaler and the receiver. Given our highly social nature, signals of prosociality are important for both the displayer and the perceiver. Angry faces, in contrast, do not have similar benefits on the receiver and signaler side. Certainly, the receiver wants to detect signals of threat, but many displays of anger do not necessarily want to be detected. If our ancestors did want to signal threat (for example, to scare off an interloper), then they likely employed other attention grabbing mechanisms like shouting, brandishing weapons, even baring teeth in an open-mouthed display (that nevertheless does not seem to yield an ASE; see Becker et al., 2011, Experiments 4b–c). There would thus be little selective pressure to make static displays of anger more salient in crowds. Dynamic expressions may thus eventually reveal a greater detection advantage for anger. We should note, however, that we have not yet seen any evidence of it. In Becker et al. (2012), we produced dynamic displays by a rapid—30 ms—flicker of facial morphs from neutral to happy to equate the onset, offset, and rate of change of the faces, and continued to see HSEs.

The visual search task thus still has potential to be used to tease apart which aspects of biological signals can grab attention. These discoveries can in turn be used to think about why we came to use certain signaling features and not others. This coevolution of human signals and their perceptual receivers is thus a potentially promising but as-yet neglected use of the visual search task.

Underexplored possibilities for "face in the crowd" research

There are still promising possibilities for threat detection research as well as for using the visual search task to explore face-in-the-crowd processing. With regard to threat, we should first acknowledge that the human mind may be found to privilege angry expressions at later stages of information processing. Indeed, there is evidence that angry face hold attention once they have been seen (e.g., Becker, Rheem, Pick, Ko, & Lafko, 2019; Belopolsky, Devue, & Theeuwes, 2011), and that they are more strongly represented in primary or working memory (Becker, Mortensen, Anderson, & Sasaki, 2014). However, happy faces *also* seem to be privileged at these stages. Future work will have to ferret out the priorities held by different layers information processing.

Even more promising is the possibility that different participants might show different profiles of prioritizing expressions of emotion. For example, participants chronically worried about interpersonal aggression might detect angry faces in the crowds more efficiently than participant who are not as preoccupied. This could happen even if these participants still show the overall happy-face detection advantage, so long as the difference between angry and happy detections can be explained by the threat vigilance. Trait anxiety, for example, appears to facilitate the detection of angry faces (Byrne & Eysenck, 1995). Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, and van IJzendoorn (2007) review a number of such effects and find that such individual difference measures often affect the earliest of attentional processing. We might also see shifts in speed–accuracy trade-offs as a function of participant level variables. If threat-vigilant participants have higher hit rates detecting angry faces, but also increased false alarms to say "anger" when no angry face is present, this reveals a bias with important implications for real-world circumstances.

Critically, however, researchers still need to be mindful of the many issues we have already discussed. For example, if one were to use schematic stimuli with potential low level confounds (e.g., eyebrows), inferences about preattentional emotion detection are unwarranted. In one study, stimuli such as these were used to show that individuals with autism spectrum disorders were as good as typical controls at detecting emotional expressions (Ashwin, Wheelwright, & Baron-Cohen, 2006). This is a surprising discovery given that emotion perception is often impaired in such individuals. However, without a complementary demonstration using a diverse set of photographic stimuli, the more parsimonious explanation is that this result does not stem from emotion

perception, but rather shows that the ability to detect angular edge-like features remains intact. A similar criticism can be leveled at another study which used these simple schematic stimuli to show that emotion perception was unimpaired in the elderly, with a similar counterexplanation—that only angular feature detectors are spared—remaining unexplored (Mather & Knight, 2006).

Despite these cautions, it is still possible to use the visual search for faces to look for participant-level variables that moderate the results. There is potential to use them as diagnostic tools, so long as we can discriminate psychological effects from stimulus-based effects. When one is contrasting the difference in the detectability of two different emotional expressions, one is comparing two different stimulus types and confounds can intrude. But when one is looking at how one group of participants performs relative to another, the exact same stimulus items and displays can be used, and any differences that are observed must be due to psychological factors.

There is precedent for such a design principle in other fields. In cognitive neuroscience, designs that seek this level of stimulus consistency conform to what is called the "Hillyard principle" (e.g., Luck, 2014). In word recognition, this design principle has been called the "ideal strategy manipulation" (Stone & Van Orden, 1993; see also Becker et al., 2006). This need not be a conscious strategy, but is rather a manipulation that changes how a participant can differentially recruit automatic and controlled information processing components to better accomplish a task. Consider, again, semantic priming (Meyer & Schvaneveldt, 1971): The same stimulus "NURSE" can be recognized, categorized, and pronounced more rapidly if it follows the prime "DOCTOR" relative to the prime "BREAD." Priming has both automatic and controlled aspects. At short prime–target stimulus onset asynchronies (SOAs), the effect seems largely automatic and is relatively unaffected by other manipulations like the proportion of related pairs (Neely, 1977). At longer prime–target SOAs, one begins to see participants employing strategies like anticipating what related words might be presented. These strategic factors are in turn moderated by other factors, like relatedness proportion, working memory loads, and so forth (Heyman, Van Rensbergen, Storms, Hutchison, & De Deyne, 2015; Hutchison, 2007). There are many demonstrations that the same stimulus displays can elicit different performance consequences because of other trials, design features, or individual differences. What is crucial is that different participants see exactly the same target and distractor displays. What varies is something else that goes on in other trials, blocks, or the broader environment. These manipulations must not affect the visual display if inferences about vision are to be drawn (e.g., sitting next to a smelly or threatening confederate would conform to this design principle; positioning that confederate in front of the display would not). These constraints

are important, but there are many possible realizations, and very few have been tried.

Conclusions

In this paper we have tried to paint a cautionary picture of the pitfalls of using Treisman's visual search methodology to look at the realistic application of finding a threat in the crowd. We have also emphasized, however, that there is the potential for positive contributions if one exercises appropriate diligence in applying these methods.

We first reviewed evidence that initial forays into the face-in-the-crowd effect that purported to show that angry faces pop out of crowds seem to be wrong. Instead, a variety of stimulus-level factors intruded to create spurious results that have driven incorrect inferences, and these effects are still being cited in ways that assume their validity. Some angry schematic faces do yield search efficiencies (some sad ones as well), but here, low-level confounds make the demonstrations inconclusive without the results of real faces to complement them.

And real angry faces also do not seem to drive efficient search. They instead slow things down when crowds are angry, which sometimes makes it appear that happy faces are detected more slowly than angry faces are, but only for the spurious reason that they are embedded in angry crowds.

In contrast, happy faces *are* efficiently detected. They are also efficiently rejected when serving as distractors. Several features of the happy facial expression seem to be at play in this search efficiency, and we discussed the possibility that some of the more vivid features may been selected because of their ability to preempt and diffuse conflict. Much more needs to be done to explore this possibility.

We do see considerable value in continuing to use the visual search to explore threat detection if stimuli are held constant and participant level variables are measured or manipulated. For example, trait anxiety does indeed seem to speed the detection of angry faces. Although it may well be the case that faces do not pop out at all, unless there is some low-level feature that drives visual search, consider the possible finding that participants efficiently find one African man in a crowd of Caucasian men. It would be silly to infer that this is caused by stereotypes about threat when a more obvious and parsimonious explanation is that color differences drive efficient visual search. But if those that score high on a measure of implicit prejudice show a bigger effect than those that score low on that measure, there would be evidence that individual differences can be leveraged to ferret out real-world consequences and establish new diagnostic tools. Other manipulations—for example, emotional priming—might have similar effects or induce speed–accuracy trade-offs that could reveal important

effects. There are thus considerable unexplored possibilities for "face in the crowd" research.

Whatever forms these future researchers take, they would be best advised to follow certain basic guidelines:

- Vary the crowd size so that search slopes can be assessed.
- Account for the speed with which distractors are rejected by considering the target-absent search rates, or ensure that all of the distractor arrays are equivalent.
- Ensure that participants are processing the stimulus signal of interest rather than low-level features that are correlated with this signal.
- Vary the distractors and targets in ways that keep participants from learning to use any low-level features to complete the task.
- Jitter the positions of the items in the crowds so that textural gestalts cannot be exploited.

We hope that this overview of face-in-the-crowd research inspires both caution and curiosity. We have tried to survey a representative sample of this research, but have certainly missed some, and we hope this has not distorted the perspective we offer. It should be clear, though, that if one attends to the lessons of Treisman's feature integration theory and the guided search models that followed, one can avoid the pitfalls that have plagued past "face in the crowd" research and explore the many new possibilities that this method still affords.

References

Ashwin, C., Wheelwright, S. & Baron-Cohen, S. (2006). Finding a face in the crowd: Testing the anger superiority effect in Asperger Syndrome. *Brain and Cognition, 61,* 78–95.

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

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S., & Neel, R. (2011). The face in the crowd effect unconfounded: Happy faces, not angry faces, are more efficiently detected in the visual search task. *Journal of Experimental Psychology: General, 140*(4), 637–59. https://doi.org/10.1037/a0024060

Becker, D. V., Goldinger, S., & Stone, G. O. (2006). Perception and recognition memory of words and werds: Two-way mirror effects. *Memory & Cognition, 34,* 1495–1511. https://doi.org/10.3758/BF03195914

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*(2), 179–190. https://doi.org/10.1037/0022-3514.92.2.179

Becker, D. V., Neel, R., Srinivasan, N., Neufeld, S., Kumar, D., & Fouse, S. (2012). The vividness of happiness in dynamic facial displays of emotion. *PLOS ONE, 7*(1), e26551. https://doi.org/10.1371/journal.pone.0026551

Becker, D. V., Mortensen, C. R., Anderson, U. S., & Sasaki, T. (2014). Out of sight but not out of mind: memory scanning is attuned to threatening faces. *Evolutionary Psychology.* https://doi.org/10.1177/147470491401200504

Becker, D. V., & Srinivasan, N. S. (2014). The vividness of the happy face. *Current Directions in Psychological Science, 23*(3), 189–194. https://doi.org/10.1177/0963721414533702

Becker, D. V., Rheem, H., Pick, C., Ko, A. & Lafko, S. (2019). Angry faces hold attention: Evidence of attentional adhesion in two paradigms. *Progress in brain research, 247,* 89–110.

Belopolsky, A. V., Devue, C., & Theeuwes, J. (2011). Angry faces hold the eyes. Visual Cognition, 19(1), 27–36. https://doi.org/10.1080/13506285.2010.536186

Byrne, A., & Eysenck, M. W. (1995). Trait anxiety, mood and threat detection. *Cognition and Emotion, 9*(6), 549–562. https://doi.org/10.1080/02699939508408982

Dimberg, U., & Öhman, A. (1996). Behold the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion, 20,* 149–182. https://doi.org/10.1007/BF02253869

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & Psychophysics, 63,* 1004–1013. https://doi.org/10.3758/BF03194519

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2003). Negative facial expression captures attention and disrupts performance. *Perception & Psychophysics, 65*(3), 352–358. https://doi.org/10.3758/bf03194566

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–29. https://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologist Press.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & Emotion, 14*(1), 61–92. https://doi.org/10.1080/026999300378996

Fox, E., Russo, R., & Dutton, K. (2002). Evidence for delayed disengagement from emotional faces. *Cognition and Emotion, 16*(3), 355–379. https://doi.org/10.1080/02699930143000527

Fridlund, A. J. (1994). "Human Facial Expression: An Evolutionary View". San Diego, CA: Academic Press, Cambridge University Press.

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin, 134*(5), 662–676. https://doi.org/10.1037/0033-2909.134.5.662

Hager, J. C., & Ekman, P. (1979). Long-distance transmission of facial affect signals. *Ethology and Sociobiology, 1*(1), 77–82. https://doi.org/10.1016/0162-3095(79)90007-4

Hahn, S., & Gronlund, S. D. (2007). Top-down guidance in visual search for facial expressions. *Psychonomic Bulletin & Review, 14*(1), 159–165. https://doi.org/10.3758/bf03194044

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917–924. https://doi.org/10.1037//0022-3514.54.6.917

Heyman, T., Van Rensbergen, B., Storms, G., Hutchison, K. A., & De Deyne, S. (2015). The Influence of Working Memory Load on Semantic Priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 911–20. https://doi.org/10.1037/xlm000005

Horstmann, G. (2007). Preattentive face processing: What do visual search experiments with schematic faces tell us? *Visual Cognition, 15*(7), 799–833. https://doi.org/10.1080/13506280600892798

Horstmann, G. (2009). Visual search for schematic affective faces: Stability and variability of search slopes with different instances.

Cognition and Emotion, 23(2), 355–379. https://doi.org/10.1080/02699930801976523

Horstmann, G., & Bauland, A. (2006). Search asymmetries with real faces: Testing the anger-superiority effect. Emotion, 6(2), 193–207. https://doi.org/10.1037/1528-3542.6.2.193

Horstmann, G., Becker, S., Bergmann, S., & Burghaus, L. (2010). A reversal of the search asymmetry favoring negative schematic faces. Visual Cognition, 18(7), 981–1016. https://doi.org/10.1080/13506280903435709

Horstmann, G., Scharlau, I., & Ansorge, U. (2006). More efficient rejection of happy than of angry face distractors in visual search. Psychonomic Bulletin & Review, 13(6), 1067–1073. https://doi.org/10.3758/bf03213927

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(4), 645–662. https://doi.org/10.1037/0278-7393.33.4.645

Luck, S. J. (2014). An introduction to the event-related potential technique. Cambridge, MA: MIT Press.

Mather, M. & Knight, M. (2006). Angry faces get noticed quickly: Threat detection is not impaired among older adults. The journals of gerontology. Series B, Psychological sciences and social sciences, 61, 54–7.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90(2), 227–234. https://doi.org/10.1037/h0031564

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 106(3), 226–254. https://doi.org/10.1037/0096-3445.106.3.226

Öhman, A., & Dimberg, U. (1978). Facial expressions as conditioned stimuli for electrodermal responses—Case of preparedness. Journal of Personality and Social Psychology, 36(11), 1251–1258. https://doi.org/10.1037/0022-3514.36.11.1251

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. Journal of Personality and Social Psychology, 80(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. Psychological Review, 108(3), 483–522. https://doi.org/10.1037/0033-295x.108.3.483

Purcell, D. G., & Stewart, A. L. (2010). Still another confounded face in the crowd. Attention, Perception, & Psychophysics, 72(8), 2115–2127. https://doi.org/10.3758/BF03196688

Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. Perception, 25(9), 1091–1108. https://doi.org/10.1068/p251091

Scherer, K., & Wallbott, H. (1994). Evidence for universality and cultural variation of differential emotional response-patterning. Journal of Personality and Social Psychology, 66(2), 310–328. https://doi.org/10.1037//0022-3514.66.2.310

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In D. V. Blake & A. M. Uttley (Eds.), Proceedings of the Symposium on Mechanisation of Thought Processes (Vol. 2, pp. 511–529). London, England: Her Majesty's Stationery Office.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 19(4), 744–774. https://doi.org/10.1037//0096-1523.19.4.744

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. Vision Research, 51(11), 1273–1278. https://doi.org/10.1016/j.visres.2011.04.002

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. Psychological Review, 114(1), 71–103. https://doi.org/10.1037/0033-295X.114.1.71

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. Cognitive Psychology, 12(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Williams, M. A., Moss, S. A., Bradshaw, J. L., & Mattingley, J. B. (2005). Look at me, I'm smiling: Visual search for threatening and non-threatening facial expressions. Visual Cognition, 12(1), 29–50.

Wolfe, J. M. (1992). The parallel guidance of visual attention. Current Directions in Psychological Science, 1(4), 124–130. https://doi.org/10.1111/1467-8721.ep10769733

Wolfe, J. M. (1998). What can 1,000,000 trials tell us about visual search? Psychological Science, 9(1). Retrieved from https://search.bwh.harvard.edu/new/pubs/MillionTrials.pdf

Wolfe, J. M. (2018). Visual search. In J. T. Wixted (Ed.), Stevens' handbook of experimental psychology and cognitive neuroscience (4th ed., pp. 569–424). Hoboken, NJ: John Wiley & Sons.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience, 5(6), 495–501. https://doi.org/10.1038/nrn1411