The face-in-the-crowd effect: Threat detection versus iso-feature suppression and collinear facilitation

Matthew J. Kennett
Centre for Sensorimotor Performance, School of Human Movement and Nutrition Sciences, University of Queensland, Queensland, Australia

Guy Wallis
Centre for Sensorimotor Performance, School of Human Movement and Nutrition Sciences, University of Queensland, Queensland, Australia

Are people biologically prepared for the rapid detection of threat posed by an angry facial expression, even when it is conveyed in the form of a schematic line drawing? Based on visual search times, the current literature would suggest that the answer is yes. But are there low-level explanations for this effect? Here, we present visual search results for schematic faces using current best practice, based on a concentric search array and set size manipulation. Using this approach, we replicate the classic search advantage for angry over happy faces. However, we also report a comparable effect when abstract plus- and square-shaped stimuli—derived from the angry and happy schematic faces respectively—are used within the same paradigm. We then go on to demonstrate that, while reduced, the effect remains after removal of the circular surround, bringing us closer to the source of the effect. We explore the possibility that the source of this search asymmetry could be the iso-feature suppression and collinear facilitation model proposed in Li's (1999a, 1999b, and 2002) bottom-up model of saliency. Simulations with this model using the abstract stimuli align with the corresponding behavioral results (i.e., the plus shape was found to be more salient than the square). Given the deliberate similarities between these abstract shapes and the respective face stimuli, we propose that the underlying cause for the asymmetries typically found using schematic faces, may be more related to early visual processing of line orientation than threat detection.

Introduction

In principle, any individual trait that confers an advantage to an organism over its competitors—be they con- or hetero-specific—is likely to be preserved through the processes of natural selection. Although the precise nature of the advantage may be subtle, traits that offer any form of basic survival advantage will presumably be among those most strongly selected for. In the case of humans, if an individual fails to live long enough to procreate and at least briefly support their young, the chances of passing on their genetic material are slim to none, irrespective of how well they may otherwise be adapted to their environmental niche. Any enhanced ability to detect signs of imminent danger that confer a critical survival advantage must surely be selected for over the generations.

For example, the automatic and immediate attraction of a person’s attention toward a shape in their visual field corresponding to a dangerous creature such as a spider or a snake, could reasonably be a trait that nature selects. The sooner someone becomes aware of such a creature in their immediate vicinity, the more able they will be to avoid becoming envenomated and potentially killed. And as one might expect, there has been experimental support found for the actual development of this trait in humans (e.g., Isbell, 2006; Ohman, 1986). However, is this just scratching the surface? Are there other dangers that might be processed especially quickly or accurately? One of the most significant threats to the survival of early man was likely not spiders or snakes but, rather, other humans. Perhaps humans have developed a trait to immediately and automatically detect threat in the appearance of a face.

The hypothesis that angry/threatening facial expressions might be afforded enhanced detection rates, was first tested by Hansen and Hansen (1988) using classical visual search (VS) techniques: The paradigm originally developed to test a visual processing theory proposed by Treisman and Galade (1980), called feature integration theory (FIT).
Using black-and-white sketches electronically rendered from photographs (Purcell, Stewart, & Skov, 1996) of people expressing anger (frowning) or happiness (smiling), Hansen and Hansen (1988) found that search (reaction) times (RTs) for identifying an angry face among happy faces were unaffected by an increase in set size, whereas search times for finding a happy face among angry ones increased. In line with the claims of FIT, they took this as evidence that angry faces contained a primitive visual feature, that could be processed in a separate early-vision subsystem, which was not present in happy faces. They went on to suggest that the particular subsystem involved in processing this feature (whatever it was), evolved due to the survival advantage this provides to the avoidance of threat, a suggestion that is beyond the claims of FIT.

There are several reasons to be skeptical about Hansen and Hansen’s (1988) angry-face search-advantage finding, as well as their conclusion that such an effect represents an evolved threat-avoidance ability. There are concerns around the veracity of FIT and the VS paradigm itself, especially considering the type of stimuli used in their study. And there is also the possibility that such an effect, even if real, simply reflects visual processes completely unrelated to the detection of threat.

Hansen and Hansen themselves, with others (Hampton, Purcell, Bersine, Hansen, & Hansen, 1989) appreciated that their earlier conclusion may have been erroneous. When they replicated their earlier experiment (Hansen & Hansen, 1988), while taking account of the target’s positioning, they found a positioning effect for both types of target. When targets were above center, detection times were faster than when below. This is not what would be expected in the case of an angry-face target “popping out,” but instead suggested that the angry-face target search advantage effect, was actually a happy-face distractor search advantage effect. Nonetheless, the authors maintained the effect was due to emotion perception.

In practice, the clear pop-out versus serial self-terminating search dichotomy envisaged by FIT was not generally borne out in VS results. This motivated development of an alternative theoretical interpretation, called guided search (Wolfe, 1994, 2007; Wolfe, Cave, & Franzel, 1989; Wolfe & Gancarz, 1996), where rather than attention being immediately drawn to a target, as in pop-out, or randomly directed around a scene to end up with steep search slope in a serial self-terminating search, different stimuli can guide attention to a target to a degree that varies along a continuum. Flat versus steep search slopes simply represent the extremes. Furthermore, it is debatable whether or not even extremely efficient searches (i.e., pop-out) are reliably diagnostic of basic features. For example, search pop-out asymmetries have been found to the advantage of backwards letters compared with their standard counterparts (e.g., Wang, Cavanagh, & Green, 1994), and upside silhouettes of animals compared with the right way up (Wolfe, 2001). It is possible that novelty may also be a basic featural dimension, although this remains a matter of debate.

It has also been shown that there are potentially confounding low-level causes for set size effects other than the presence or absence of basic features. For example, increasing display eccentricity causes a set size effect, due to differences in retinal sensitivity, and so too does display density, due to the effects of lateral masking or crowding (Carrasco, Evert, Chang, & Katz, 1995). A manipulation of set size is unavoidably accompanied by either one of these. It has also been suggested that set size effects can largely be explained as due to decision processes within a signal detection framework (e.g., Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, 1995; Palmer, Verghese, & Pavel, 2000), as opposed to the perceptual-attentional processes proposed by FIT.

But perhaps the greatest problem with Hansen and Hansen’s (1988) study, which initiated a branch of research called the face-in-the-crowd-effect (FICE) that continues to this day, relates to their use of photographs to produce VS stimuli. Photographic pictures are far more complicated than the type of stimuli typically used in VS experiments, since this complexity makes it difficult to determine which parts of the stimuli are the cause for any effect. In Hansen and Hansen’s case, their effort to try to make the photographic images simpler, by rendering them into one-bit (i.e., each pixel either black or white) sketches, introduced confounding artefactual differences (dark and light patches) that were likely the cause for the effect. Hampton et al. (1989) removed the dark patch (post rendering) from the images, and in terms of the overall effect, only found an angry-face search advantage when the set size was nine, but not four. Purcell et al. (1996) controlled for the artefacts by rendering the photos into grayscale images (with 256 gray levels) to directly test if these might have been the actual cause for Hansen and Hansen’s original result. They failed to find any effect.

The unintentional introduction of salient dark and/or light patches via image manipulation is not the only issue with using photos. There are many potential confounds. For example, a light patch in the exposure of teeth might be present in either a happy or angry face but not the other (Horstman, Lipp, & Becker, 2012), depending on which facial expression database is used (e.g., the database of Ekman & Friesen, 1976; or The Karolinska Directed Emotional Faces database, or Lundqvist, Flykt, & Öhman, 1998), or which particular pictures are used from any particular set. Failure to find consistent results in replications of the original
study is likely due to numerous low-level confounds afflicting the use of real pictures of faces (Ohman, Juth, & Lundqvist, 2010; Ohman, Lundqvist, & Esteves, 2001), which motivated a move to using pictorially simpler line-drawing images, in the same vein as early object recognition tasks aimed at controlling for the influence of texture, color, and other extraneous cues (Biedermann, 1987).

### Controlling for photo confounds: Schematic faces

Biederman (1987) discovered in his research on object recognition, that when trying to elucidate what might constitute a basic visual component, line drawings offer a more controlled set of stimuli than pictures of real objects. Faces come in line drawings too, known as emoticons or smiley faces, or schematics. FICE researchers plied their trade with variations of these forms, and the results here were more consistent (Horstmann, 2007), with negative faces almost always enjoying a search advantage effect. However, now it is not just angry faces eliciting the effect; sad and scheming faces do too (Nothdurft, 1993; Tipples, Atkinson, & Young, 2002; White, 1995), even though the idea that a sad face indicates threat seems more than a little odd.

However, despite their overall consistency in terms of direction, the results did exhibit considerable variability in the particular search times, which would not be expected if the angry/sad faces communicated threat equally (Horstmann, 2007, 2009). Different studies have used different versions of the paradigm as well (e.g., different set sizes, array configurations, presentation lengths, etc.), which could account for the variation. However, when Horstmann included all the different stimuli used in some of the most highly cited papers (Eastwood, Smilek, & Merickle, 2001; Fox et al., 2000; Nothdurft, 1993; Ohman, Lundqvist, & Esteves, 2001; White, 1995), within the one experimental paradigm, the variation between different stimuli remained, and the finding for a negative-face-as-target search advantage remained consistent.

The move from pictures of real faces to schematic line drawings of faces undeniably reduced the potential for (nonemotional) featural confounds; however, it is impossible to remove them altogether. The problem is that the effect might be traced back to differences in the arrangement of the component elements accompanying each different schematic face, rather than their emotional expression. Investigations which have tested hybrid, intermediate versions of happy and angry faces (i.e., “scheming,” with an up-turned mouth but down-turned eyebrows, and “worried” with a down-turned mouth but up-turned eyebrows), have found intermediate search advantage results (e.g., Ohman et al., 2001; Tipples et al., 2002).

In order to determine if nonexpression-related featural differences, resulting from the necessarily different arrangement of the line components, was the underlying cause for the schematic angry face effect, Coelho, Cloete, and Wallis (2010) created a new set of abstract stimuli. These stimuli consisted of either four radial lines projecting from the near center to the near edge of the surrounding circle, similar to the angry schematic face. Or four (approximately) concentric lines, similar to those used in the happy schematic face. Maintaining these common internal characteristics between the schematic faces and the abstract shape analogues, but rotating them by 45°, resulted in abstract shapes reportedly neutral in valence and not explicitly associated with any facial emotion (see Figure 1).

Coelho et al. (2010) found the usual search advantage for angry face targets over the happy faces; however, they also found the same effect for the corresponding, nonemotion-conveying (i.e., 45° rotated) abstract stimuli. Their reasoning for the effect was that perhaps T-junctions formed by the radial lines interacting with the circular surround were the basic feature aiding its detection. T-junctions were proposed by Julesz (1981) as potential basic features in his studies in texture segregation, where multiples of each feature are grouped (creating textures), and the ease of

<table>
<thead>
<tr>
<th></th>
<th>Face</th>
<th>Abstracted</th>
<th>Rotated 45°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td><img src="image1" alt="Happy Face" /></td>
<td><img src="image2" alt="Abstracted Face" /></td>
<td><img src="image3" alt="Rotated Face" /></td>
</tr>
<tr>
<td>Angry</td>
<td><img src="image4" alt="Angry Face" /></td>
<td><img src="image5" alt="Abstracted Face" /></td>
<td><img src="image6" alt="Rotated Face" /></td>
</tr>
</tbody>
</table>

Figure 1. Examples of the stimuli originally developed by Coelho et al. (2010) and also used in the experiments described in this article.
discriminating their boundary is tested. Features of
textures that seem to “pop-out” (i.e., boundaries very
easily discriminated), which Julesz called “textons,”
correspond to the basic features of VS.

In Coelho et al.’s (2010) study, the search advantage
effect was taken as a pure RT difference, in search
times for the different targets within a fixed number of
distractors (i.e., set size was constant). However, a
number of factors other than faster target/discrepancy
detection could be the cause for such results; for
example, differences in response production after
having detected the target/discrepancy, or variations in
stimulus encoding (Frischen, Eastwood, & Smilek,
2008). Comparing the effects of set size manipulation
avoids these issues and is thus considered a more
principled way to investigate any pre-attentive pro-
cesses in the actual search for discrepant images. To
this end, one of the aims in the present investigation is
to replicate Coelho’s experiments, using their schematic
images for both happy and angry faces as well as the
rotated abstract stimuli (i.e., encircled plus and square
shapes), but this time manipulating set size. A finding
of a similar result (i.e., a search advantage for both the
angry-face and plus-shape stimuli over their counter-
parts, but this time as search asymmetries by using the
appropriate VS methodology) may go some way
toward ending the threat versus low-level feature
debate, at least within the realm of schematic face
drawings, which itself was devised as an attempt to
circumvent confounds present when using photos of
real faces.

If there really is a threat detector for face emotion,
then this must be additional to, or maybe even
disruptive of, any existing lower level feature effect.
There should be a difference between schematic face VS
results and their abstract analogues. However, based on
the findings of Coelho et al. (2010), we do not expect
there to be any such differences, even with our revised
methodology. It is thus hypothesized that both
schematic face and abstract shape-opposing stimuli will
elicit a similar pattern of search results, with search
times being less negatively affected by an increase in set
size, for both the angry face schematic and encircled
plus-shape abstraction than their respective counter-
parts.

Experiment 1

General procedure and materials

The experiment was conducted using a PC, situated
in a quiet, darkened room. MATLAB (MathWorks,
Natick, MA) was selected for its availability, wide-
ranging use, search task compatible psychophysics
toolbox (Psychtoolbox), as well as its ability to
accurately and precisely take measurements of RTs,
isolated from interference of other running processes
(Cunningham & Walleraven, 2012). An existing Psy-
chtoolbox example experiment, designed to test peo-
lies’ memories of previously presented pictures
(“OldNewRecognitionExp”), was adapted for the
creation of a VS task experiment.

The experiment consisted of a series of static circular
arrays of schematic faces presented to participants, for
5 s each or until a response was made, via a 40-cm
(diagonally; 9:6 aspect ratio) Sony Trinitron CRT
display. A circular array, rather than the rectangular
grid pattern used in Coelho et al.’s (2010) original
study, was chosen in order to avoid any potential
effects of the target centrality (Purcell & Stewart, 2010),
and to try to enforce actual target singleton image
detection rather than allow for texture segmentation.
The display of each array was preceded by the
presentation of a fixation dot for 2 s; however, fixation
on this dot was not enforced and eye movements were
not tracked at any time during the experiment.

Each schematic face was a 256 × 256 pixel bitmap
image created using Microsoft PowerPoint from a file
provided by Coelho et al. (2010). The faces themselves
were centered within the bitmaps, such that faces with
circular surrounds had diameters of 200 pixels. The
lines making up the faces were six pixels in width.

Set sizes of the arrays were two, four, or eight faces.
For a set size of two, one face was positioned
uppermost (0°), and another face was positioned at
180°. For a set size of four, two additional faces were
positioned at 90° and 270°, respectively. For a set size of
eight, four additional faces were positioned at 45°, 135°,
225°, and 315°, respectively.

Faces with circular surrounds extended 2 cm across
the display area, and the diameter of the array, from
outermost edge to outermost edge, was 11.5 cm, such
that given a viewing distance of 60 cm, the entire array
subtended approximately 11°, and each individual
image 2°, on each participant’s eyes.

Each array (trial) consisted of either all the same face
image (same), or one face was different from the other/s
different). The position of the different face was
randomly determined. There were an equal amount of
same trials as there were different trials within each
block of trials, and there were equal amounts of each
set size. Thus, each block consisted of 12 trials as
follows:

- 2 × Set Size 2 Different Trials (happy different and
angry different were indistinguishable);
- 2 × Set Size 2 Same Trials (1 × all happy; 1 × all
angry);
- 2 × Set Size 4 Different Trials (1 × happy different;
and angry different).
• 2 × Set Size 4 Same Trials (1 × all happy; 1 × all angry);
• 2 × Set Size 8 Different Trials (1 × happy different; 1 × angry different);
• 2 × Set Size 8 Same Trials (1 × all happy; 1 × all angry);

Each participant repeated 15 blocks of these 12 trials, for both schematic face stimuli and abstract analogues (30 blocks overall), with the stimulus type order counterbalanced across participants.

Participants were informed that there would be a series of arrays displayed (following a presentation of a fixation dot), which would be either a repeat of the same image, or there would be one image different from the other/s, and for each array they were required to indicate which was the case as fast and accurately as possible, by pressing either the “1” key if they were the same, or the “2” key if different.

The first of each of the 15 blocks was used as training for the task, and was thus not included in the analysis. RTs for all other correct responses (accuracy was 91% overall) were then used as the main measure in the analysis, which was conducted using MATLAB.

Eight participants were used in each experiment, which was calculated as one more than the minimum number required for an effect size of 0.8, corresponding to that found by Coelho et al. (2010) using the same stimuli, and correlation between measures of 0.8, for a 95% confidence interval (CI), which was determined to be seven participants for a 99% assurance (Maxwell, Kelley, & Rausch, 2008). An even number of participants was required to balance order effects given it was a repeated-measures design.

Participants

Ethics approval was obtained from the University of Queensland, School of Human Movement Studies’ Ethical Review Committee. Participation was on an informed consent basis. All participants had normal or corrected-to-normal vision. Participants were recruited from the undergraduate population of the University’s School of Human Movement and Nutrition Sciences.

Conditions and participants

In Experiment 1, each face, or abstract analogue, included a circular surround. Eight participants (five females and three males), aged 18 to 51 years (M = 24.6 years) completed the experiment as described above.

Results

The initial plan was to test for differences in search slopes; however, due to nonlinearity (R²: M = 0.44 for angry-face targets, M = 0.70 for happy-face targets, M = 0.57 for plus-shape targets, and M = 0.81 for square-shape targets), a two-way repeated-measures analysis of variance (ANOVA) was conducted instead, with set size (two, four, and eight) and target type (happy/angry face and plus/square shape) as the independent variables, and RTs as the dependent variable. This analysis revealed significant interactions for participants to correctly respond to a discrepancy, due to either a single happy-face or a single angry-face image present among one, three, or seven of the opposite; F(2, 14) = 8.01, p = 0.005, with an effect-size (ηp²) of 0.53 (0.27 to 0.83, 95% CI range), as well as for when the images were abstract analogs (plus and square shapes) of these schematic faces; F(2, 14) = 7.10, p = 0.007, with an effect size (ηp²) of 0.50 ([0.27, 0.78] 95% CI).

Paired t tests found a significant difference in discrepancy detection caused by either a happy-face or angry-face singleton, when there were seven distractors (Set Size 8), but not when there were three (Set Size 4) or one (Set Size 2). For Set Size 8, angry-face singletons were found significantly faster (M = 987 ms; SD = 369 ms) than happy-face singletons (M = 1,164 ms; SD = 420 ms); t(7) = −4.12, p = 0.005, with an effect size (Hedge’s g) of −0.42 ([-0.78, −0.12] 95% CI). For Set Size 4, angry-face singletons were found faster (M = 1,035 ms; SD = 362 ms) than happy-face singletons (M = 1,091 ms; SD = 381 ms), but not significantly so; t(7) = −1.95, p = 0.093, with an effect size (Hedge’s g) of −0.14 ([−0.34, 0.03] 95% CI). And for Set Size 2, happy-face singletons were found faster (M = 875 ms; SD = 195 ms) than angry-face singletons (M = 906 ms; SD = 271 ms), but this difference was likely just due to chance; t(7) = 1.04, p = 0.331 with an effect size (Hedge’s g) of 0.13 ([−0.15, 0.42] 95% CI), which is to be expected, since there was no actual difference in the stimuli for the two target-present conditions when there were only two images present in total.

Paired t tests found a similar pattern of results for simple effects with the abstract shapes. For Set Size 8, plus-shape singletons were found significantly faster (M = 844 ms; SD = 146 ms) than square-shape singletons (M = 1,037 ms; SD = 297 ms); t(7) = −2.94, p = 0.022, with an effect size (Hedge’s g) of −0.77 ([−1.55, −0.11] 95% CI). For Set Size 4, plus-shape singletons were found slightly and nonsignificantly slower (M = 892 ms; SD = 215 ms) than square-shape singletons (M = 871 ms; SD = 227 ms); t(7) = 0.61, p = 0.560, with an effect size (Hedge’s g) of 0.09 ([−0.24, 0.43] 95% CI). And for Set Size 2, discrepancy detection was slightly faster when the plus-shape appeared as the target (M = 987 ms; SD = 369 ms) than when the square-shape did (M =
987 ms; SD = 369 ms); \( t(7) = -0.62, p = 0.553 \) with an effect size (Hedge’s \( g \)) of \(-0.11 ([-0.51, 0.28] 95\% \) CI), even though there was no actual difference between the two stimuli, as reflected in the chance-like difference result.

The type of VS task adopted for these experiments, whereby participants were allowed plenty of time (up to 5 s) to search and respond for each stimulus presentation—as opposed to the other type where stimuli are presented only briefly (less than 1 s)—uses RTs as the dependent variable. The results in this regard have been reported above. However, to ensure there was no speed–accuracy trade-offs in the participants’ responses, we also analyzed the results in terms of accuracy (i.e., in terms of correct vs. incorrect responses). This accuracy analysis suggests there were no speed–accuracy trade-offs, with the differences between the happy/angry and plus/square target conditions in terms of accuracy, being in the opposite direction to those in terms of RTs. The accuracy interaction for faces was at chance level; \( F(2, 14) = 0.17, p = 0.847 \), with an effect size \( (\eta_p^2) \) of 0.02 ([0.00, 0.52] 95\% CI). It was, once again, not significant for the abstract shapes, although it did approach significance in this case; \( F(2, 14) = 3.53, p = 0.057 \), with an effect size \( (\eta_p^2) \) of 0.34 ([0.17, 0.67] 95\% CI). All results for Experiment 1, including accuracy simple effects, are provided in Figure 2.

Discussion

As predicted, searching for an angry face among happy face distractors was, overall, significantly more efficient than the reverse, consistent with the angry-face hypothesis, although this was only the case for the larger set size of eight. However, as also predicted, abstract shape images derived from the angry and happy schematic faces elicited a similar pattern of effects consistent with the hypothesis that, in both cases, this is due to lower level featural differences in the schematic drawings interacting with normal vision processes. This suggests that communication of threat is an unnecessary further explanation for this effect. It should be noted that response accuracy in this experiment was lower than usual for typical VS experiments (Wolfe, 1998); however, such experiments typically involve simpler stimuli that are more easily distinguished. There is a considerable number of nondiagnostic features shared by the angry and happy faces with surrounds (surround, eyes, nose) that may well have led to error rates rising, as well as the RTs being slower than typically found with VS experiments. Note that all stimuli were presented at the same eccentricity, with an angular range of 3.5° to 5.5°. The fact that they lay outside the central fovea may also have contributed to the increased latency (Carrasco et al., 1995; Wolfe, O’Neill, & Bennett, 1998). The RTs for our experiment were, however, comparable to Coelho et al.’s (2010) results that used identical stimuli. Also, importantly, as is clear from the accuracy and mean RT set size results, the pattern of accuracy changes over set sizes corroborates the RT findings (i.e., there is no evidence of a speed–accuracy trade-off).

Our results support those obtained by Coelho et al. (2010), which used the same stimuli, this time incorporating the paradigmatically correct manipulation of set size to determine the relative search efficiencies, while using the same symmetrical design with current best practice improvements. This suggests that their results were likely not due to any response biases and that other potential stimulus-based confounds were sufficiently controlled. However, one thing that remains unresolved is the source for the search asymmetry. This effect, which is strong, has been replicated across many labs and studies.

As mentioned earlier, Coelho et al. (2010) suggested the underlying cause for the effect might be due to the “T junction”-like parts of the angry face where its mouth and eyebrows approach the outer circular line, and which are absent in both the happy face and the square/diamond abstracted shape counterparts. There are, however, still reasons for querying this. While T junctions were proposed by Julesz (1981) to be basic feature, a more recent, comprehensive study by Wolfe and Horowitz (2004) listed it as a probable non-attribute.

The T junction idea stemmed from the previous FICE literature that consistently pointed to the important role of the surround. Of particular relevance, Dickins and Lipp (2014) tested for search differences using the Coelho et al. (2010) abstract stimuli, both with and without the circular surround, albeit while repeating Coelho et al.’s omission of a set size manipulation in their experimental methodology. They too found a consistent advantage for the cross/plus shape over the diamond/square shape, whether in original face-like orientation (cross vs. diamond) or the neutral-valence eliciting 45° rotated orientation (plus vs. square); but this difference was only significant with the surrounds. Without the surrounds, search times for targets of both shapes was much faster, but the differences were not statistically significant. Dickins and Lipp concluded that the cause for the significant effect found with the surrounds is most likely due to some kind of interaction between the inner parts of these images with the circular surround. That said, they did admit to the possibility that the overall decrease in RTs seen for faces without a surround led to the possibility of a floor effect in their data. We also noted that many of the original reports used faces in rigid
rectangular arrays, allowing for the possibility that the surrounds were needed to aid segmentation in that special case.

Craig, Becker, and Lipp (2014) performed a conjunction search investigation, comparing schematic happy and angry faces as targets among a common mixture of intermediate faces. Their distractors were either sad faces, with the happy face down-turned eyebrows and the angry face down-turned mouth, or “scheming” faces, with the angry face up-turned eyebrows and the happy-face up-turned mouth. They found no effect of set size, in terms of a significant difference between slopes, whether the surrounds were included or removed. They did, however, find a significant difference in the overall RT means, both with and without surrounds in favor of the happy face,
although it was unclear whether the cause was emotional or due to a low-level feature. They did not include nonface, but rather face-derived, abstract shapes in their study. Horstmann, Becker, Bergmann, and Berghaus (2010) and Becker, Horstmann, and Remington (2011) also found a reversal of the search advantage between happy and angry schematic faces in a symmetrical design. By changing the shape of the surround from a circle, to a three-quarter moon shape, such that the bottom of the circle (the “chin”) was angled inwards toward the mouth, they were able to show that happy “dented” faces enjoyed a search advantage over their angry “dented” face counterparts.

Given the uncertainty around the importance of the surrounds’ involvement in the effect found in our first experiment, we felt the logical next step was to repeat Experiment 1, using the same stimuli, but with the circular surrounds removed. If the effect is due to a low-level feature that involves the surround, be it a T junction (Coelho et al., 2010), inner–outer line conformance (Purcell & Stewart, 2010), or perceptual grouping (Becker et al., 2011), it should no longer be evident with the surrounds removed. However, if the effect is due to a low-level feature within the inner part of the stimuli, the effect should remain, and be evident in both the facial stimuli as well as the abstracted shapes. Purcell and Stewart (2010) did also find a (smaller) search advantage for angry schematic faces with the surrounds removed in their experiments, concluding from this that perhaps there is an emotion perception contribution to the effect independent of low-level feature causes. A finding of the effect only for schematic faces, with surrounds removed, but not the abstract analogues, would support this conclusion.

Experiment 2

Participants and conditions

Conditions for Experiment 2 were exactly the same as for Experiment 1, except the circular surrounds were removed from all stimuli. Participants for Experiment 2 consisted of eight paid volunteers in total: six males and two females aged 18 to 33 (M = 22.5 years).

Results

Search slopes for responses with no-surround stimuli were also nonlinear (R²: M = 0.64 for angry-face targets, M = 0.83 for happy-face targets, M = 0.42 for plus-shape targets, and M = 0.55 for square-shape targets), so a two-way repeated-measures ANOVA was conducted with these instead as well. Removing the surrounds reduced both the significance and sizes of the interaction effects for the schematic face stimuli; F(2, 14) = 3.64, p = 0.053, with an effect size ($\eta^2_p$) of 0.34 ([0.10, 0.73] 95% CI), and this reduction was closely matched with the abstract shape stimuli; F(2, 14) = 2.96, p = 0.085, with an effect size ($\eta^2_p$) of 0.30 ([0.09, 0.63] 95% CI).

The simple effects for differences between target types without surrounds, at each set size, also mirrored those for with surrounds, but with the differences reduced. For Set Size 8, angry-face singletons were found significantly faster (M = 922 ms; SD = 219 ms) than happy-face singletons (M = 1,080 ms; SD = 222 ms); t(7) = −3.08, p = 0.018, with an effect size (Hedge’s g) of −0.66 ([−1.31, −0.12] 95% CI). For Set Size 4, angry-face singletons were found faster (M = 978 ms; SD = 173 ms) than happy-face singletons (M = 1,004 ms; SD = 237 ms), but not significantly so; t(7) = −0.65, $p = 0.537$, with an effect size (Hedge’s g) of −0.12 ([−0.53, 0.28] 95% CI). And for Set Size 2, happy-face singletons were found slightly faster (M = 880 ms; SD = 144 ms) than angry-face singletons (M = 929 ms; SD = 144 ms); t(7) = 0.91, p = 0.394 with an effect size (Hedge’s g) of 0.31 ([−0.46, 1.13] 95% CI).

Similar results to the faces stimuli were again found in the simple effects for the abstract shapes, without surrounds. For Set Size 8, plus-shape singletons were found significantly faster (M = 646 ms; SD = 110 ms) than square-shape singletons (M = 684 ms; SD = 120 ms); t(7) = −3.18, p = 0.015, with an effect size (Hedge’s g) of −0.31 ([−0.61, −0.05] 95% CI). For Set Size 4, plus-shape singletons were found very slightly faster (M = 655 ms; SD = 166 ms) than square-shape singletons (M = 663 ms; SD = 135 ms); t(7) = −0.29, p = 0.778, with an effect size (Hedge’s g) of −0.04 ([−0.39, 0.29] 95% CI). And for Set Size 2, discrepancy detection was slightly faster when the square shape was programmatically inserted as the target (M = 601 ms; SD = 101 ms) than when the plus-shape was (M = 641 ms; SD = 151 ms); t(7) = 1.22, p = 0.261 with an effect size (Hedge’s g) of 0.29 ([−0.24, 0.86] 95% CI), even though there was no actual difference between the two stimuli, as reflected in the chance-like difference result.

Accuracy with surrounds removed was very high overall (≥90%), and there was no evidence of a speed–accuracy trade-off. The accuracy interaction for faces was approaching significance; F(2, 14) = 0.31, p = 0.74, with an effect size ($\eta^2_p$) of 0.31 ([0.02, 0.80] 95% CI), however, it was near chance level for the abstract shapes; F(2, 14) = 1.53, p = 0.251, with an effect size ($\eta^2_p$) of 0.18 ([0.01, 0.70] 95% CI). All results for Experiment 2, including accuracy simple effects, are provided in Figure 3.
Discussion

Although reduced to marginal significance, with the surrounds removed, there was still a target type by set size interaction favoring an angry-face search advantage effect. However, as with Experiment 1 (i.e., stimuli with circular surrounds), in Experiment 2 it was also found that the difference was only evident in the results for Set Size 8. Importantly though the same was also the case for the abstract stimuli. This suggests that there is at least some contribution to the effect other than what might be caused by interactions between the
inner and outer line components that are present with surrounds, as suggested by Coelho et al. (2010), Purcell and Stewart (2010), and Dickins and Lipp (2014), for example. It should also be noted that the removal of the surrounds greatly improved the overall accuracy rate of Experiment 2 compared with Experiment 1, even though the experiments were conducted in the same manner on a similar student population. There are two plausible mechanisms at play here. First, the surround may act as an interfering (crowding) feature (e.g., Levi, 2008; Whitney & Levi, 2011; Xu, Liu, Dayan, & Qian, 2012), making it harder for participants to identify the different faces or shapes. Or, the same effect may be due to there simply being more overlap of (hence) nondiagnostic featural elements as was discussed above.

The fact that the effect using the schematic faces was repeated when using the abstract analogues, both in terms of the effect itself as well as the reduction when the surrounds were removed, suggests that the effect is not due to the detection of threat, but rather, to some form of basic visual feature present within both kinds of stimuli. Overall search times were also quicker with stimuli that had fewer overlapping features (i.e., abstract vs. face, and no-surround vs. surround) presumably because it is easier to detect a discrepancy when there are fewer or no similarities. This may have been the cause for the reduction in the effect when surrounds were removed. Perhaps a further increase in set size, or more data, is required in this case to properly reveal the effect.

Given all of these considerations, the results are consistent with the hypothesis that radial lines enjoy a search advantage over concentrically arranged features. The rest of this article is dedicated to searching for a model of visual saliency that is consistent with this asymmetry.

**Simulations**

If it is indeed the case that the search advantage found for both the schematic angry face and its derivative abstract plus shape, over their happy-face/square-shape counterparts, is due simply to differences in the early processing of the common visual components contained within the opposing images, then this should be predicted by computational simulations of such processes. There are several related computational models of visual processing, using only the information present in the stimuli, to determine which specific location/s within a scene would draw the focus of attention, based on the theoretical work of Koch and Ullman (1985). The basic premise of these models is that contains the greatest difference between its features and the features of closely surrounding areas along one or more common featural dimension—is the most attractive region for attention, and thus this is where the focus of attention is drawn in a “winner-takes-all” fashion. Once this region is visited, it is then discounted as an area of further interest and attention is drawn to the next most salient location, and so on, until all relatively salient regions in the scene have been attended to.

Thus, the results found in our and many of the other VS experiments on schematic faces, may be replicated by these models if angry-face and plus-shape images are considered computationally more salient than their happy-face counterparts.

**Itti and Koch’s saliency map**

One of the most widely used models incorporating the Koch and Ullman (1985) approach is Itti and colleagues’ (Itti & Koch, 2001; Itti, Koch, & Niebur, 1998) saliency map, which is available for free download as a MATLAB implementation as the Saliency Toolbox (Walther & Koch, 2006). This model (see figure 4 from Itti & Koch, 2001) can take any image file as its input, and produce a topographic map of that file where each region is assigned a scalar value, which represents its saliency (i.e., a saliency map). This is achieved by first extracting information about different feature elements (color, luminance, and orientation) over several different spatial scales through linear filtering, and creating feature maps from this information, which are then subjected to iterative normalizing and within-feature competition processes, before finally being combined into a single representation of overall salience, the saliency map.

To see if this model could explain our results, we created image files for each target-present condition by taking a screenshot of each stimulus for when the target appeared at the 12 o’clock position (this, or the 6 o’clock position where the only positions where it could be kept the same for each set size), which we used as input for the Saliency Toolbox MATLAB implementation, using the toolbox’s default parameters (Walther & Koch, 2006). The initial result for each input file was a saliency map, as well as a selected area overlaid on top of the original image, as the most salient to which it would be expected that attention would be directed. We then continued through the simulation, which subsequently applied an “inhibition of return” factor to the selected area, before creating a new saliency map and selecting the region with the next highest saliency value. We continued with this process until two locations had been selected that contained two different image items.
(i.e., the target and a distractor). The results of this process for each image file are shown in Appendix 1.

Having attained an expected number of item visits to detect discrepant items for each of the target-present conditions, we then incorporated Wolfe’s (1994) expected timings for given number of item searches to allow us to compare the simulation result expectations with our own experimental results. Wolfe proposes an expected RT for typical VS performance of 400 ms plus 50 ms for each item visit, normally distributed with a standard deviation of 25 ms multiplied by the square root of the number of item visits. Table 1 shows the resulting simulation RT results in comparison to the corresponding experimental results, and Figure 4 provides graphical comparisons.

Discussion

The Itti and Koch (2001) simulation results accurately predicted the difference and direction for the schematic face stimuli, both with and without surrounds, and even predicted there would only be a difference at Set size 8 (but not four), albeit only with the surrounds. However, the simulations predicted a difference in the opposite direction, both in regard to our experimental results and the schematic faces, when it came to the derivative abstract shape stimuli. And this was again the case for both with and without the surrounds. According to the simulations, a discrepancy should have been detected faster, at both set sizes, and been less affected by additional distractors, than when the square shape was the singleton.

These simulation results suggest that the attempt to capture the important featural components of the schematic face stimuli in the abstract shape stimuli, but in a way that does not depict a face, may not have been successful. Perhaps the similar experimental results using these two stimulus types were due to higher level, top-down influences rather than lower level, bottom-up influences. However, another cause for the simulation results may be just that the Itti and Koch (2001) model, as implemented in the MATLAB-based Saliency Toolbox using default parameters, is not the appropriate model to use for our stimuli. Li’s (1999a, 1999b, and 2002) V1 Saliency Hypothesis model, however, can make analytical predictions based only the differences in the line orientations of the inner components.

Li’s V1 saliency hypothesis

Whereas Koch and Ullman (1985) models (e.g., Itti & Koch, 2001) envisage a master saliency map that exists downstream of the primary visual cortex (V1), Li (e.g., 1999a, 1999b, and 2002) has proposed a model in
which this saliency map emerges directly in the output of V1 itself. This is called the V1 saliency hypothesis, which, using a model that simulates V1 responses, has been shown to account for a number of classical VS data. Of relevance here is that the model also provides a theoretical explanation for why radial lines might be more salient than concentric lines, and hence easier to detect.

The V1 saliency hypothesis (Li, 1999a, 1999b, and 2002) proposes that horizontal interactions between neurons in Layers 2 to 3 of V1 (which necessarily exhibit proximal receptive fields) are the origin of many, if not all, visual saliency effects. When a stimulus’s feature value (e.g., orientation, color, etc.) matches the preferred feature value of a V1 neuron in the feature dimension that the neuron is tuned, its activation is high. But each neuron also receives contextual influence from its neighboring neurons, which can be inhibitory or excitatory depending on the preferred features of the interacting neurons and spatial relationship between the receptive fields.

For example, two nearby neurons tuned to the same orientation will facilitate each other if they are displaced along the same direction as their preferred orientation (e.g., tuned to a vertical orientation and displaced vertically), but will inhibit each other if they are displaced orthogonally to this orientation (e.g., tuned to vertical, but displaced horizontally). The latter is referred to as iso-feature (orientation) suppression, and the former as collinear facilitation. This contextual influence decays with distance between the two receptive fields, and the interaction pattern is a bowtie–like association field (Field, Hayes, & Hess, 1993; Li, 1998). Collinear suppression and iso-orientation suppression both appear to be applicable to the stimuli used in our experiments. The vertical bars of the square shape are displaced horizontally (and the horizontal bars displaced vertically), and thus neurons tuned to these

<table>
<thead>
<tr>
<th>Set sizes</th>
<th>Visits (objects)</th>
<th>Sim. RT (ms)</th>
<th>Act. RT (ms)</th>
<th>Diff. (Act. – Sim.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>500</td>
<td>947</td>
<td>447</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>763</td>
<td>263</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>904</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>621</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>621</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>621</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>621</td>
<td>404</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set sizes</th>
<th>Visits (objects)</th>
<th>Sim. RT (ms)</th>
<th>Act. RT (ms)</th>
<th>Diff. (Act. – Sim.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>700</td>
<td>1,035</td>
<td>335</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>750</td>
<td>1,091</td>
<td>341</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>700</td>
<td>892</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>550</td>
<td>871</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>978</td>
<td>478</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>600</td>
<td>1,004</td>
<td>404</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>600</td>
<td>655</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>663</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>663</td>
<td>163</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set sizes</th>
<th>Visits (objects)</th>
<th>Sim. RT (ms)</th>
<th>Act. RT (ms)</th>
<th>Diff. (Act. – Sim.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>9</td>
<td>850</td>
<td>1,087</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1,200</td>
<td>1,164</td>
<td>–36</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1,150</td>
<td>844</td>
<td>–306</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>800</td>
<td>1,037</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>550</td>
<td>922</td>
<td>372</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1,000</td>
<td>1,080</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>800</td>
<td>646</td>
<td>–154</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500</td>
<td>684</td>
<td>184</td>
</tr>
</tbody>
</table>

Table 1. Simulated versus actual results comparisons. Notes: Sim. = simulated; Act. = actual; Diff. = difference; SS = set size. See Appendix 1 for how simulated visit results were determined. Simulated reaction-times (RT) calculated in accordance with Wolfe (1994).
orientations at the respective retinal locations stimulated should inhibit each other’s activity according to Li’s model. Conversely, the vertical bars of the plus shape are displaced vertically (and horizontal bars displaced horizontally), meaning that the contextual influence from the vertically (or horizontally) tuned neurons stimulated by the presence of these bars, should be co-facilitative. Therefore, the plus-shape (and angry-face) stimuli should be more salient than the square-shape (and happy-face) stimuli, and this should guide a more efficient search for the plus-shape/angry-face stimuli than the reverse. Simulation of the V1 model of contextual influences with the original model parameters (Li, 1998; Li, 1999b) confirmed these expectations (see Appendix 2). The results (summarized in Table 2) are evidently in general agreement with our experimental results.

It should be noted that if the stimuli include a circular surround, Li’s (1999a, 1999b, and 2002) model would regard them as highly salient due to collinear facilitation (also referred to as contour enhancement). However, this salience is equal between the two opposing stimuli, and the overall difference in salience is still determined by the inner line arrangements. This offers an explanation as to why we find comparable effects both with and without surrounds, an outcome which the Itti and Koch (2001) model fails to predict.

A final and important note to add at this point is that we have omitted one seemingly crucial test of Li’s (1999a, 1999b, and 2002) model, namely a test of the schematic faces themselves. Unfortunately, Li created her model some years ago and it is not currently programmatically able to simulate processing with sufficient resolution to include the detail of eyes or the curved contours of a mouth. As such it is not possible to draw any firm conclusions as to whether her model can reproduce all of the effects described in this paper. Hence, we offer her model as one possible avenue to an explanation that appears more theoretically promising than other classical models.

### General discussion

Unlike their real-face counterparts, schematic-face images have produced an almost unanimously consistent set of results regarding the threat-detection superiority hypothesis of Ohman, Hansen, and others. However, this consistency should not necessarily be taken to imply the absence of a confound. There may still be a low-level, nonemotion-based cause for the angry schematic face to draw attention more readily than a happy schematic face.

Coelho et al.’s (2010) finding (replicated by Dickins & Lipp, 2014) of a similar search advantage for nonface images that share many features of the original faces, but none of the associated valence, provide support for this idea. That said, the fact that they did not attempt a set manipulation or control for stimulus eccentricity limits the extent to which the results can be related to the broader VS literature. In Experiment 1 we reused

<table>
<thead>
<tr>
<th>Target type</th>
<th>Surround</th>
<th>Scale</th>
<th>Set size</th>
<th>Max target&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Max distractor&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Average&lt;sup&gt;c&lt;/sup&gt;</th>
<th>z score&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plus</td>
<td>No</td>
<td>Fine</td>
<td>4</td>
<td>0.61</td>
<td>0.52</td>
<td>0.53</td>
<td>2.15</td>
</tr>
<tr>
<td>Square</td>
<td>No</td>
<td>Fine</td>
<td>4</td>
<td>0.52</td>
<td>0.60</td>
<td>0.57</td>
<td>-1.50</td>
</tr>
<tr>
<td>Plus</td>
<td>No</td>
<td>Fine</td>
<td>8</td>
<td>0.61</td>
<td>0.52</td>
<td>0.52</td>
<td>3.19</td>
</tr>
<tr>
<td>Square</td>
<td>No</td>
<td>Fine</td>
<td>8</td>
<td>0.52</td>
<td>0.60</td>
<td>0.58</td>
<td>-2.27</td>
</tr>
<tr>
<td>Plus</td>
<td>Yes</td>
<td>Fine</td>
<td>4</td>
<td>0.58</td>
<td>0.57</td>
<td>0.50</td>
<td>1.08</td>
</tr>
<tr>
<td>Square</td>
<td>Yes</td>
<td>Fine</td>
<td>4</td>
<td>0.57</td>
<td>0.58</td>
<td>0.52</td>
<td>0.90</td>
</tr>
<tr>
<td>Plus</td>
<td>Yes</td>
<td>Fine</td>
<td>8</td>
<td>0.58</td>
<td>0.57</td>
<td>0.49</td>
<td>1.10</td>
</tr>
<tr>
<td>Square</td>
<td>Yes</td>
<td>Fine</td>
<td>8</td>
<td>0.57</td>
<td>0.58</td>
<td>0.53</td>
<td>0.92</td>
</tr>
<tr>
<td>Plus</td>
<td>No</td>
<td>Gross</td>
<td>4</td>
<td>0.56</td>
<td>0.48</td>
<td>0.49</td>
<td>1.88</td>
</tr>
<tr>
<td>Square</td>
<td>No</td>
<td>Gross</td>
<td>4</td>
<td>0.48</td>
<td>0.55</td>
<td>0.53</td>
<td>-1.51</td>
</tr>
<tr>
<td>Plus</td>
<td>No</td>
<td>Gross</td>
<td>8</td>
<td>0.58</td>
<td>0.49</td>
<td>0.47</td>
<td>2.94</td>
</tr>
<tr>
<td>Square</td>
<td>No</td>
<td>Gross</td>
<td>8</td>
<td>0.50</td>
<td>0.58</td>
<td>0.53</td>
<td>-0.54</td>
</tr>
<tr>
<td>Plus</td>
<td>Yes</td>
<td>Gross&lt;sup&gt;g&lt;/sup&gt;</td>
<td>4</td>
<td>0.45</td>
<td>0.47</td>
<td>0.51</td>
<td>0.71</td>
</tr>
<tr>
<td>Square</td>
<td>Yes</td>
<td>Gross&lt;sup&gt;g&lt;/sup&gt;</td>
<td>4</td>
<td>0.47</td>
<td>0.45</td>
<td>0.42</td>
<td>1.34</td>
</tr>
<tr>
<td>Plus</td>
<td>Yes</td>
<td>Gross&lt;sup&gt;g&lt;/sup&gt;</td>
<td>8</td>
<td>0.59</td>
<td>0.61</td>
<td>0.43</td>
<td>1.84</td>
</tr>
<tr>
<td>Square</td>
<td>Yes</td>
<td>Gross&lt;sup&gt;g&lt;/sup&gt;</td>
<td>8</td>
<td>0.61</td>
<td>0.59</td>
<td>0.43</td>
<td>2.29</td>
</tr>
</tbody>
</table>

Table 2. Summary of Li Model simulations. Notes: aResponse to the most salient bar in the target. bResponse to the most salient bar in individual non–target (distractor) items. cAverage response to individual input bars. dz score of the most salient bar in the target. eNote that (a) the circular surround in each item at this scale is aliased by the model’s sampling such that it is not symmetrical with respect to the square and the cross. In particular, each sampling gap in the circle misses a bar that should be parallel to the nearest square bar inside the circle, and thus should exert more iso-orientation suppression on the square bar if this bar in the gap was properly sampled. And that (b) the saliency order between the square and the cross is opposite from expected. It is likely that (a) caused (b).
their stimuli but employed a more standard circular search array, as well as manipulating the size of the array.

The common wisdom at the time of Coelho et al.’s (2010) publication was that the face-in-the-crowd effect disappears if the surround is removed (Purcell & Stewart, 2005; Schübö, Gendolla, Meinecke, & Abele, 2006). This led the authors to speculate as to why the surround was so important. They suggested that the radial lines of the angry-face, cross-, or plus-shaped stimuli might interact with the surround to produce a T-shaped junction and that this might serve as a basic search feature. However, in the interim, Dickens and Lipp (2014), reported that although removing the surrounds from the abstract shapes reduced the search advantage for the angry faces, the predicted difference was still evident (albeit not statistically significant). In this current paper we sought to investigate this further. In Experiment 2 we repeated Coelho et al.’s experiment but now with the surrounds removed. We reasoned that if the effect was indeed due to interactions between inner features and the surround, then the effect should be lost after the removal of the surrounds. What we found for the face stimuli was generally in line with what Dickens and Lipp reported. Removal of the surrounds reduced the size of the angry face search advantage but did not negate it entirely; indeed, in our case the difference remained statistically significant at a set size of eight. The pattern for the abstract stimuli was similar, with a significant search advantage once again emerging for a set size of eight but not for the other set sizes. Overall RTs were remarkably quick, especially for the smaller set sizes and especially for the abstract stimuli. In fact, the response times approached the limits of a simple RT task, suggesting the discrimination part of the task had become almost trivial at smaller set sizes. The fact that the search times for the faces were slower than for the abstract shapes might be explained by the interference caused by the remaining (nondiagnostic) features common to these images, namely the “nose” and “eyes.”

If, as our results suggest, the surrounds are not crucial to the effect, what might be the reason for the search advantage? In the final phase of the article we applied two models of visual saliency to our stimuli so see if the trends we saw in our behavioral results could be replicated. We first tested Itti and Koch’s (2001) saliency map model (Saliency Toolbox MATLAB implementation using default parameters). This model predicted the direction of our results for the schematic face stimuli, both with and without surrounds, but predicted the opposite direction to our results for the abstract versions, again both with and without surrounds. The second model that we tested was Li’s (1999a, 1999b, and 2002) V1 saliency hypothesis, and this was shown to explain all of the effects associated with the abstract stimuli tested here. Unfortunately, as described above, because we were unable to test the schematic faces, our assertion that her model could likewise explain results for the face stimuli remains only speculation at this stage. Note, though, that the collinear facilitation at the heart of Li’s model is a mechanism that is present neurophysiologically in V1 but absent from Itti and Koch’s saliency map model. If V1’s response does encapsulate at least some aspects of saliency, it provides a reason to expect, a priori, a greater level of saliency for the angry-face schematic faces, and its abstract derivative, than their counterparts.

We should note that the specific V1 circuit model that Li developed offers only a rough quantitative model of V1. This is not only because of our limited knowledge of the quantitative parameters defining the operation of V1, but also because the V1 model only operates at one spatial scale rather than at multiple scales (due to Li’s original intent to focus only on the most essential characters of contextual influence in her model). Hence, we should take the quantitative results of the V1 model simulation only as a reference. Nonetheless, as we noted in the main text, the nature of the plus-shape advantage in evoking higher V1 responses rests qualitatively, rather than quantitatively, on the well-accepted finding of the nonisotropic nature of the contextual influences between V1 cells, and is thus insensitive to quantitative details. It is worth adding that we retained the parameters of the V1 model originally used nearly 20 years ago (Li, 1998; 1999a), demonstrating the robustness of the original model. Note also that the V1 model that we used does not capture the activity of V1 neurons that are not tuned to orientation. These neurons are tuned to isotropic circular shapes, akin to the center-surround profiles preferred by retinal ganglion cells, except bigger. These neurons are part of the multiscale receptive fields in V1 (see Zhaoping, 2014 for details), and would likely respond to the circular surround of the facial inputs. It is likely that these neurons contribute to the saliency of faces, although a priori there seems to be no reason for these neurons to prefer angry or happy faces.

The VS literature has revealed that there are a great variety of image pairs that can elicit search asymmetries (Wolfe, 1998, 2001; Wolfe & Horowitz, 2004), and some search asymmetries have their basis in neural mechanisms beyond V1 (Zhaoping & Frith, 2011), so it is also possible that the consistent result that we found for asymmetries in favor of the angry face schematic and its abstract derivative relied on different and additional causes. Perhaps the angry face was found faster due to a special mechanism that evolved to better protect us from the potential threat such faces warn us of, as suggested by Hansen and Hansen (1988), while the plus shape was found faster due to iso-feature suppression and collinear facilitation. However, that
there is a common cause, consistent with the latter, strikes us as the more parsimonious explanation.

One potential challenge to our conclusions lies in studies that have demonstrated the singular importance of the facial surrounds when they are distorted to match or contrast with the inner features. It might not appear immediately obvious how we can reconcile our results with those findings (e.g., Becker et al., 2011; Horstmann et al., 2010). In practice though, their results may offer further support for the conclusions we draw here. On inspection of their stimuli, it appears that altering the surrounds in the way that they did led to interactions between the features and surround consistent with Li’s model. Indeed Becker et al. (2011) raised this as a possibility in their report. In the end, it remains possible that the search advantage effects found throughout our experiments and that are consistent with a common cause—for example, the colinear facilitation and iso-feature suppression of inner line segments—may in fact each have different causal mechanisms. Future research may discover a means of distinguishing between the possible causes of the search asymmetries described, but until then, we would argue that parsimony favors a common cause.

Keywords: face-in-the-crowd, visual search, salience, attention, primary visual cortex (V1)

Acknowledgments

We would like to thank Zhoaping Li for explaining her model and conducting simulations with it based on the stimuli used in our experiments (see Appendix 2).

Commercial relationships: none.
Corresponding author: Matthew J. Kennett.
Email: m.kennett@uq.edu.au.
Address: Centre for Sensorimotor Performance, School of Human Movement and Nutrition Sciences, University of Queensland, Queensland, Australia.

References


**Appendix 1: Itti and Koch saliency map simulations**

**Method**

- MATLAB Saliency Toolbox (Walther & Koch 2006) was downloaded from http://www.saliencytoolbox.net/.
- Implemented in MATLAB version R2015a.
- Image files were created by running each experimental condition with each target image located at a 12 o’clock position, taking a screenshot 540 pixels square, and saving as a .png image file.
- Each file was then used as the input for the runSaliency.m Saliency Toolbox program, using default parameters, as per defaultSaliencyParams.m that is called by runSaliency.m.
- The (winner take all) saliency map output file was saved for the first run—i.e., before inhibition of return was applied (applyIOR.m)—before proceeding with subsequent runs until two different objects had been calculated as containing the winning location. These locations were indicated by yellow or green (for the first) outlines, connected by red lines between subsequent locations, overlaid on top of the input image files.
- Both the number of fixations (F)—that is, the raw number of locations whether or not within separate objects—and number of objects (O) were counted from the first location until the last.

**Results**

Figures A1 through A4.
Figure A1. (A) Angry face, with surrounds. (B) Happy face, with surrounds.
Figure A2. (A) Plus shape, with surrounds. (B) Square shape, with surrounds.
Figure A3. (A) Angry face, without surrounds. (B) Happy face, without surrounds.
Figure A4. (A) Plus shape, without surrounds. (B) Square shape, without surrounds.
Appendix 2: Li V1 saliency hypothesis simulations

Method

Simulation of Li’s V1 model was done using the original model parameters published in Li (1998) and Li (1999b). Each input shape (e.g., plus shape or square shape, with or without the outer circle [surround]) is sampled or seen by the underlying V1 model neurons as a collection of individual bar segments. This is because each model neuron is tuned to a particular visual location $i$ and a particular orientation $\theta$, and hence its optimal input is a bar segment $(i\theta)$ at this particular visual location (receptive field) and of this orientation. Li’s model is such that the centers $i$ of the receptive fields are discrete grid points of a regular grid (orthogonal or hexagonal grid). Hence, any part of the visual input not falling close enough to any of the discrete grid points is not optimally sampled by this model and would cause aliasing. This is seen in the gaps in the sampling of the circle shape (for the facial surround) by Li’s model in the following simulations.

Li’s model has only a single spatial scale (i.e., sizes of the receptive fields or the distance between neighboring grid locations $i$). Hence, input images can be scaled up or down to fit this model scale. Two different scales (“fine” and “gross”) were used for the simulation for comparison. The gross scale is the simplest representation of the experimental images, but introduces aliasing effects that bias and are a likely cause for a reversal of the results compared to expectations, when including the surrounds. This aliasing effect is reduced in with fine scale input images, and results conform to expectations for these. All the sampled bar segments $(i\theta)$ are assumed to have the same, intermediate level, input contrast of $I_i^\theta$ in Li’s model (Li, 1998, 1999).

The results are shown as the input image (seen as the collection of bars $(i\theta)$ from the perspectives of the model neurons) and the saliency map. The saliency map shows the proxy saliency value $S_i$ at each grid location $i$, defined as the maximum response of the V1 model neurons whose receptive fields share the same grid location. (Note that the real saliency value is the proxy value $S_i$ relative to the proxy saliency values at other visual locations; see Zhaoping, 2014, since saliency of a location is a measure of this location to attract attention relative to other visual locations that could also attract attention.)

More specifically, $S_i = \max_\theta |g_\theta(x_i)|$, in which $g_\theta(x_i)$ is the (temporally averaged) response of a model pyramidal cell whose receptive field is at $i$ and prefers orientation $\theta$. The response levels $g_\theta(x_i)$ of the V1 model neurons are within the range $[0, 1]$ by the original model design. We quantify the real saliency of location $i$ using a $z$ score (Li, 1999), defined as $z_i = \frac{S_i - \bar{S}}{\sigma_S}$, in which $\bar{S}$ and $\sigma_S$, respectively, are the average and standard deviation of $S_i$ over all grid locations $i$ having nonzero inputs in the input image.

Results

Input and saliency heat-map images (Figures A5 through A8).

Figure A5. Fine scale, no surrounds.
Figure A6. Fine scale, with surrounds.

Figure A7. Gross scale, no surrounds.
Figure A8. Gross scale, with surrounds.