

SCIENTIFIC REPORTS



OPEN

Crowding for faces is determined by visual (not holistic) similarity: Evidence from judgements of eye position

Alexandra V. Kalpadakis-Smith¹, Valérie Goffaux^{2,3,4} & John A. Greenwood¹ 

Crowding (the disruption of object recognition in clutter) presents the fundamental limitation on peripheral vision. For simple objects, crowding is strong when target/flanker elements are similar and weak when they differ – a selectivity for target-flanker similarity. In contrast, the identification of upright holistically-processed face stimuli is more strongly impaired by upright than inverted flankers, whereas inverted face-targets are impaired by both – a pattern attributed to an additional stage of crowding selective for “holistic similarity” between faces. We propose instead that crowding is selective for target-flanker similarity in all stimuli, but that this selectivity is obscured by task difficulty with inverted face-targets. Using judgements of horizontal eye-position that are minimally affected by inversion, we find that crowding is strong when target-flanker orientations match and weak when they differ for *both* upright and inverted face-targets. By increasing task difficulty, we show that this selectivity for target-flanker similarity is obscured even for upright face-targets. We further demonstrate that this selectivity follows differences in the spatial order of facial features, rather than “holistic similarity” *per se*. There is consequently no need to invoke a distinct stage of holistic crowding for faces – crowding is selective for target-flanker similarity, even with complex stimuli such as faces.

We have all experienced the difficulty of trying to locate a friend in a busy shopping centre or our favourite shirt in a pile of clothes. This task becomes especially hard when the face or item of clothing that we are searching for is located not in our central/foveal vision, but our visual periphery. Although part of this difficulty can be attributed to reductions in peripheral visual acuity^{1,2}, peripheral vision is further limited by *crowding* – the curious phenomenon whereby objects that are otherwise readily identifiable in isolation become jumbled and indistinguishable in clutter^{3,4}. The strength and spatial scale of crowding^{5,6} has led to it being described as the fundamental limitation on object recognition in peripheral vision⁷.

Crowding disrupts the identification of a range of fundamental visual dimensions, including orientation^{6,8–10}, colour^{11–13}, size¹⁴, and motion^{13,15}. Crowding is also selective for differences between the target object and surrounding flanker elements along these dimensions. For simple stimuli, crowding is strong when features in the target and flanker elements are similar in these dimensions, and weak when they differ^{10,11,13,16–18}. Importantly, this selectivity for target-flanker similarity is typically *symmetric*: a red target surrounded by green flankers will reduce crowding as much as a green target amongst red flankers.

The generality of this selectivity for target-flanker similarity has been questioned by the observation that the crowding of faces is qualitatively different from crowding for simple objects. Louie *et al.*¹⁹ found that the identification of an upright target face was more strongly disrupted by upright than inverted flankers, whereas the identification of an inverted target face was equally impaired by *both* upright and inverted flankers. This *asymmetry* was attributed to a selectivity for “holistic similarity”, a criterion that is met when both target and flanker faces are upright, and thus processed holistically^{20–24}. Cropped images of houses were also found to be equally impaired by upright and inverted house flankers, matching the pattern seen with inverted target faces. From these findings,

¹Experimental Psychology, University College London, London, United Kingdom. ²Research Institute for Psychological Science, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. ³Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. ⁴Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands. Correspondence and requests for materials should be addressed to A.V.K.-S. (email: alexandra.smith.13@ucl.ac.uk)

Louie *et al.*¹⁹ claim that similarity in orientation modulates crowding *only* when the target is processed holistically and not for objects processed in a part-based or featural manner (i.e. houses and inverted faces). Although it is odd that this latter “featural” pattern of crowding differs from the above studies with simple stimuli^{10,11,13,16} (where a symmetric selectivity for target-flanker similarity is observed), Louie *et al.* concluded that there are distinct crowding mechanisms for holistically processed faces and featurally processed objects. These findings have been further taken to suggest that crowding consists of multiple independent stages in the visual hierarchy^{7,19,25–27}.

We propose an alternative explanation. Although the inversion of a face clearly changes its potential for holistic processing^{20–23}, it also increases the difficulty with which the identity of that face is recognised. We suggest that this increase in task difficulty could obscure the more general selectivity for target-flanker similarity that is typically observed for crowding. In other words, it may be these variations in task difficulty, and not the presence or lack of holistic processing *per se*, that cause the curious asymmetric selectivity in the crowding of faces. A demonstration of this nature would eliminate the need for additional processing stages to account for the crowding of faces, and suggest that common principles apply for the crowding of both simple and complex objects. Here, we begin by replicating the “holistic similarity” effect, before introducing an eye-judgement task²⁸ that allows the measurement of face crowding with matched difficulty across upright and inverted target conditions, as well as the independent manipulation of task difficulty. We argue that there is indeed a common selectivity for target-flanker similarity that determines the crowding of both simple objects and complex elements such as faces.

Experiment 1

Our first aim was to replicate the findings of Louie *et al.*¹⁹ with some variations in stimuli and experimental design, in order to verify that any conflicting findings were not due to differences in these factors. We thus began with an identity-matching task where observers judged whether a peripheral target face had the same or a different identity as a reference face (Fig. 1A). The target face was presented either upright or inverted, in isolation (“uncrowded”) as well as in the presence of flanker faces (“crowded”) that themselves could be either upright or inverted.

Methods. *Observers.* Seven observers participated in Experiment 1 (5 females, $M_{\text{age}} = 23.9$ years). This sample size was deemed sufficient given that it doubles that of prior studies¹⁹. Two of the authors (AKS and JAG) took part in all experiments; the rest of the observers were naïve. All had normal or corrected-to-normal vision. For all experiments, experimental protocols and procedure were approved by the University College London Research Ethics Committee. The protocols and procedure complied with the Declaration of Helsinki, and written informed consent was obtained from all participants prior to each experiment.

Apparatus. Experiments were programmed with MATLAB (MathWorks) on a Macintosh computer running PsychToolbox^{29–31}. Stimuli were presented on a 21" cathode ray tube monitor (LaCie Electron 22 Blue IV) with a resolution of 1152×870 pixels and a 75 Hz refresh rate. The monitor was calibrated with a Minolta CS100A photometer to give mean and maximum luminance values of 50 cd/m² and 100 cd/m². Stimuli were viewed binocularly from 50 cm, with head movements minimised using head and chin rests. Experiments took place in a dark room, with responses made using a keyboard.

Stimuli. 14 adult female Caucasian faces with a neutral frontal gaze were selected from the Radboud Faces Database³². Using Adobe Photoshop, all faces were grayscale filtered, and edited to fit their internal features inside an oval window of 272×395 pixels. A generic background face texture was created, which was free of external cues like glasses and the hairline that can contribute to identification accuracy^{33,34}. The features of each face (eyes, nose, mouth, and chin) were placed onto this matched background, along with a generic pair of eyebrows that was created for all faces. Luminance values were adjusted to span the range from 0–100 cd/m² in all faces. Face RaDF-090-12³² was selected as the reference face for all experiments. The remaining 13 faces were used as distractor targets and flankers (Fig. 1B). When flankers were present, six flankers were presented on each trial, fixed on an ellipse around the target with a position along the ellipse that varied from trial to trial (see Fig. 1B).

The size of the faces was determined individually for each observer during practice blocks. By varying the size of face stimuli parametrically we selected a size where observers were capable of at least 90% accuracy with uncrowded upright targets and between 70–80% for targets surrounded by matched flankers. By using such a high performance criterion to determine stimulus size, it is unlikely that any limitations from within-face crowding³⁵ (crowding between the features of the face) play a substantial role in our results. Sizes ranged from 2.4–3.6° horizontally and 3.5–5.2° vertically, with means of 3.1° and 4.5°, respectively.

The relative centre-to-centre distance between target and flankers was fixed at $1 \times$ the width of the target. This resulted in variation in the absolute centre-to-centre target-flanker separation between observers, given the above size variations. Nonetheless, if we assume that the region of interference is half the eccentricity of the target, following standard measurements^{4,5}, then all of our stimuli fell well within the 6° interference zone for crowding at 12° eccentricity.

Procedures. Observers were required to fixate at all times on a Gaussian fixation point with an SD of 3.6 arcmins. Each trial began with the presentation of the reference face at fixation for 500 ms. The location of the reference face was jittered by 25 pixels on each trial to avoid alignment with the peripheral target that could provide extraneous cues to the task, particularly for the eye position judgements in subsequent experiments. This was followed by the presentation of the target face at 12° eccentricity in the right visual field. The target was either presented in isolation or surrounded by 6 flankers. A 1/f noise mask was then presented for 200 ms at the target location (see Fig. 1A), with a size that covered both target and flanker elements for each observer. After presentation of the mask, observers made a same/different judgement for the target face in relation to the reference face,

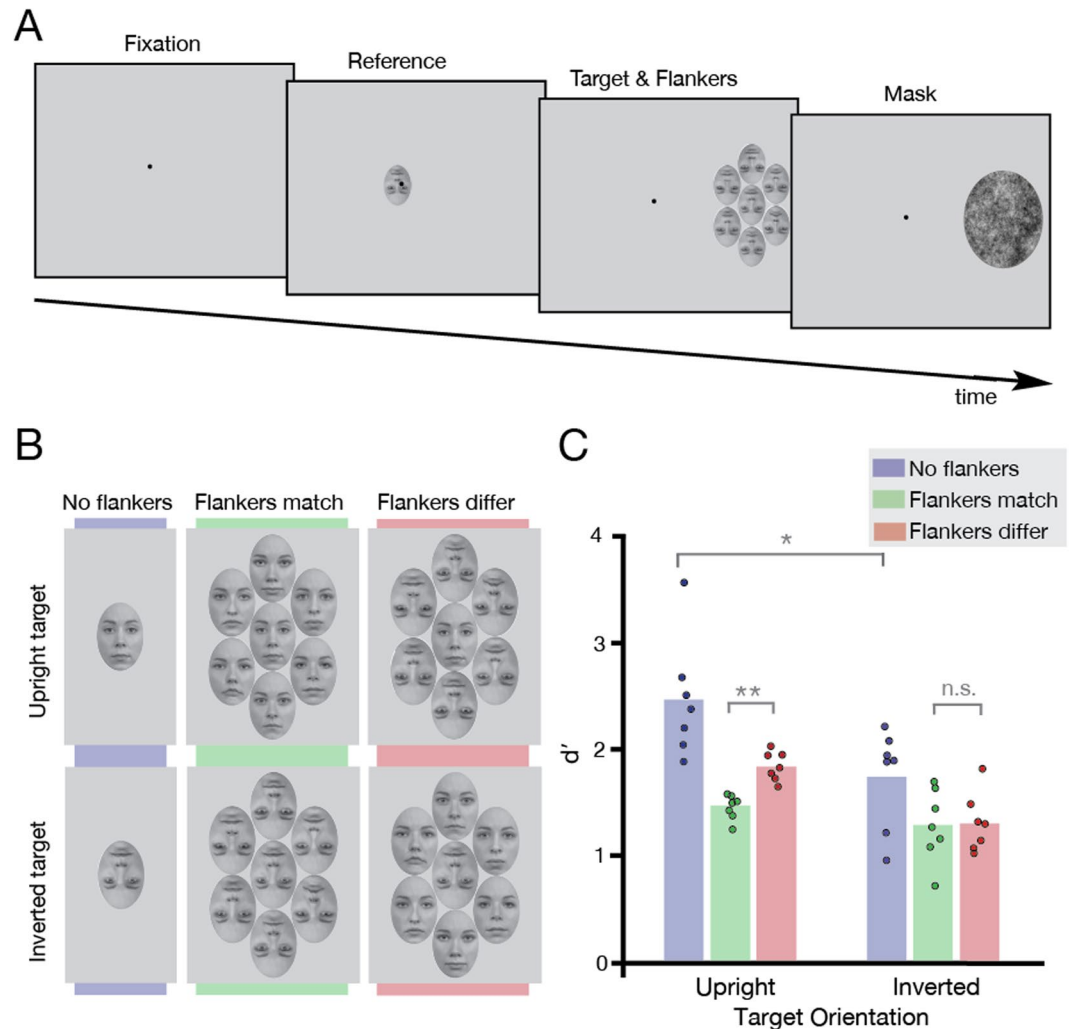


Figure 1. Stimulus time-course, design, and results of the identity-matching procedure in Experiment 1. (A) *Stimulus time course:* Each trial began with the presentation of a Gaussian fixation dot, followed by the reference face (500 ms) at fixation, then the target (500 ms, surrounded by flankers) at 12 degrees in the right visual field, and a mask (250 ms). Observers responded after the mask. A flankers-match condition with an inverted target is depicted. This time course was the same for all experiments. (B) *Experimental conditions:* Two target orientation conditions were presented in Experiment 1: upright (top panel) and inverted (bottom panel). There were three crowding conditions: the target face presented in isolation (no flankers), the target surrounded by flanker faces of the same orientation (flankers match), and the target surrounded by flankers of different orientation (flankers differ). Face stimuli were taken from the Radboud Faces Database³² and edited as described in the Methods. (C) *Results of Experiment 1:* Results plotted for the three crowding conditions, for upright and inverted targets ($N = 7$). Average d' values are depicted by bars; dots show d' for each observer. Grey brackets indicate the significance level of each condition comparison (n.s. not significant, $*P < 0.05$, $**P < 0.01$).

regarding *identity* (“is the target face the same or a different identity from the reference?”). The identity of the target matched that of the reference face on 50% of the trials, and in the rest it was one of the 13 distractor identities.

There were two target orientation conditions, upright and inverted, and three crowding conditions: “uncrowded”, where the target was presented in isolation, “crowded-same”, where the flankers matched the orientation of the target (e.g. upright flankers around an upright target), and “crowded-different”, where the orientation of the flankers differed from the target. This resulted in 2340 trials (3 crowding conditions \times 2 same/different conditions \times 13 different distractor face identities \times 5 repetitions per face identity \times 3 blocks \times 2 target orientations).

Prior to completing the set of experimental blocks for each experiment, observers completed one or more practice sessions, with auditory feedback for incorrect responses. Practice began with uncrowded trials and progressed to crowded trials after achieving at least 90% accuracy.

Results and Discussion. We examined the recognition of identity in upright and inverted target faces in three crowding conditions: the target was either uncrowded and presented in isolation, surrounded by flankers of matching orientation, or surrounded by flankers of different orientation (see Fig. 1B). For each condition, d' was calculated as an indicator of face-recognition sensitivity (see Fig. 1C) by subtracting the z-score of the false

alarm rate (incorrect “different” responses when target and reference faces were matched) from the z-score of the hit rate (correct “different” responses when target and reference faces differed)³⁶. A two-way repeated measures ANOVA was conducted with target orientation and crowding condition as factors. This analysis revealed significant main effects of target orientation ($F[1, 6] = 12.84, P = 0.012$), with identity more accurately recognised in upright than inverted faces, and crowding condition ($F[1.15, 6.91] = 19.81, P = 0.003$, Greenhouse-Geisser corrected). The interaction between target orientation and crowding condition was also significant ($F[2, 12] = 9.68, P = 0.003$). Planned comparisons using paired t-tests with Bonferroni correction showed that for upright target faces, crowding was modulated by the orientation of the flankers: performance was strongly impaired when the upright target was surrounded by similarly upright flankers (“flankers match”) and significantly less impaired with inverted flankers (“flankers differ”; $t[6] = -6.91, P = 0.0009$). Importantly, this was not the case for inverted target faces: both inverted and upright flankers were equally disruptive on performance, with no significant difference between the two ($t[6] = 0.27, P > 0.99$).

We therefore replicate the effects of Louie *et al.*¹⁹: for an identity-matching task, crowding is modulated by the orientation of the flankers for upright target faces, but not for inverted targets. The lack of selectivity for the orientation of the flankers with inverted target faces has been used to argue that crowding is selective for “holistic similarity”^{19,25}. Louie *et al.*¹⁹ further observed a trend ($P = 0.055$) towards upright flankers being more disruptive on an inverted target than inverted flankers, and speculated that this may be due to holistic information in the upright flankers over-riding the featural information of the inverted targets. Although a subset of our observers also show this pattern, an equal number of observers show the opposite (more crowding with inverted than upright flankers on an inverted target). Therefore, with a larger sample, we find no evidence to support stronger interference with upright flanker faces when the target is inverted.

The basis of the argument for holistic crowding derives from the inversion effect, a difference in the difficulty with which upright and inverted faces are recognised. This difference in difficulty has been taken to suggest distinct processing styles: holistic or configural for upright faces, and featural or part-based for inverted^{20–24}. The effect of inversion on face recognition can be seen in the current experiment by comparing upright and inverted uncrowded target face conditions in Fig. 1C. However, this leads to an alternative interpretation of the effects of crowding on faces: we propose that it is the difference in task difficulty with upright vs. inverted target faces that is responsible for the asymmetric pattern of crowding for faces, and not the differential engagement of holistic processing. Namely, the increased difficulty in recognising inverted faces could make it more difficult to then release that target from crowding with flankers of a different orientation, resulting in an apparent lack of target-flanker selectivity. If this were the case, it should be possible to reveal this selectivity for target-flanker similarity with *both* upright and inverted target faces by matching the difficulty of target recognition in upright and inverted conditions.

Experiment 2

Here we investigated the selectivity of face crowding with matched task difficulty between upright and inverted target orientations. To achieve this, we examined crowding using judgements of the horizontal separation between the eyes (interocular distance) in face stimuli, a task that has been shown to be minimally affected by inversion²⁸, and therefore of similar difficulty for upright and inverted targets. For this task, the holistic similarity account¹⁹ would predict that even with matched difficulty, crowding for inverted target faces should not be modulated by the orientation of the flankers, as inverted faces are not processed holistically. In fact, it could be argued that this eye-judgement task relies primarily on featural rather than holistic processes, given that a significant difference in the difficulty between upright and inverted face recognition is the typical marker of holistic engagement^{20,21,23,24}. Were this the case, the holistic account¹⁹ would predict that there should be no modulation of crowding by the orientation of the flankers for *either* upright or inverted target faces (i.e. the pattern observed above with inverted target faces and houses in an identity-matching task). In contrast, we propose that crowding has a general symmetric selectivity for target-flanker similarity regardless of the degree of holistic engagement, but that this can be obscured by heightened task difficulty. In this case, by matching task difficulty for upright and inverted targets, we should observe selectivity for target-flanker similarity in both target conditions.

Methods. Methods were generally similar to Experiment 1, with the following exceptions. Seven observers (4 females, $M_{\text{age}} = 25.1$ years) participated in Experiment 2, all with normal or corrected-to-normal vision.

To manipulate the horizontal eye positions within faces, the eyes and eyebrows of the reference face were shifted inwards from their original positions within the oval window by 20 pixels. This face with inward-shifted eyes served as the target in the trials when the target differed from the reference face, which was again face RaDF-090-12³² with the eyes in their original positions. For the flankers, three had no eye shift (“eyes normal”), while the rest had the eyes shifted 20 pixels inwards from their original positions (see Fig. 2A). Stimulus sizes varied between 2–4° horizontally with a mean of 2.9°, and 2.9–5.8° vertically with a mean of 4.2°. Given these variations in the sizes of the presented faces, this resulted in horizontal eye shifts between 0.15–0.30° for each observer, with a mean of 0.21°.

Observers were shown peripheral target faces at 12° eccentricity with interocular separations that were either unaltered or shifted inwards. They were required to make a same/different judgement for the target in relation to the reference face presented foveally, regarding the *horizontal eye position* (“are the eyes of the target face the same or different from the reference?”). The target had matched eye position to the reference on 50% of the trials, and in the rest it had the eyes shifted inwards.

As in Experiment 1, there were two target orientation conditions, upright and inverted, and three crowding conditions: “uncrowded”, where the target was presented in isolation, “crowded-same”, where the flankers matched the orientation of the target (e.g. upright flankers around an upright target), and “crowded-different”,

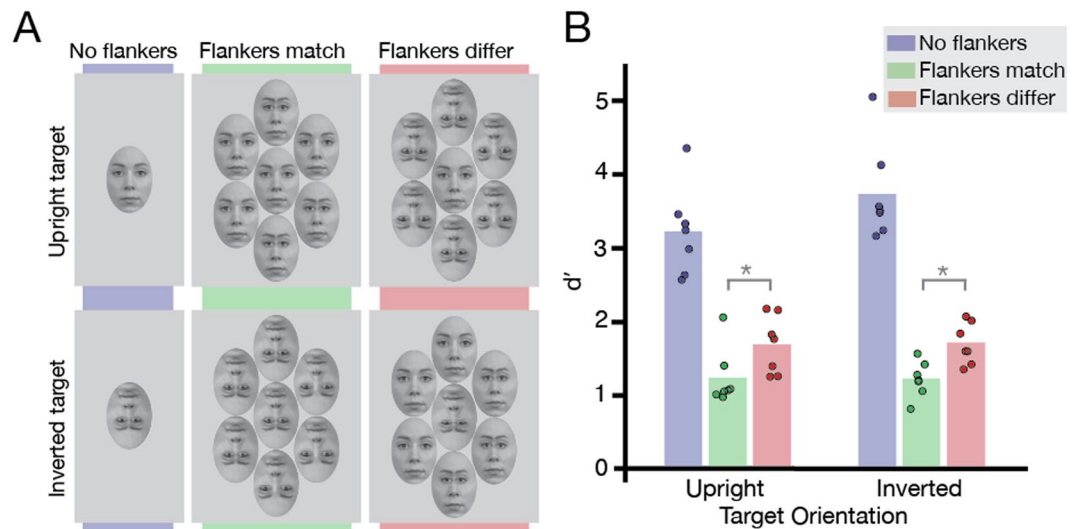


Figure 2. Design and results of the horizontal eye-shift procedure in Experiment 2. (A) *Experimental conditions:* Two target orientation conditions were presented in Experiment 2: upright (top panel) and inverted (bottom panel). There were three crowding conditions: the target face presented in isolation (no flankers), the target surrounded by flanker faces of the same orientation (flankers match), and the target surrounded by flankers of different orientation (flankers differ). The horizontal eye position of the target was either the same (“normal”, no eye-shift) or different (eyes shifted inwards). Face stimuli were taken from the Radboud Faces Database³² and edited. (B) *Results of Experiment 2:* Results plotted for the three crowding conditions, for upright and inverted targets (N = 7). Bars indicate average d' ; dots indicate d' for each observer. Grey brackets indicate the significance level of each condition comparison (n.s. not significant, * $P < 0.05$).

where the orientation of the flankers differed from the target (see Fig. 2A). This resulted in 1440 trials (3 crowding conditions \times 2 same/different conditions \times 30 repetitions per face \times 4 blocks \times 2 target orientations).

Results and Discussion. We calculated d' as a measure of observers’ sensitivity to differences in eye-separation, with results shown in Fig. 2B. A two-way repeated measures ANOVA was conducted, with target orientation and crowding condition as factors. The main effect of target orientation was not significant ($F[1,6] = 2.05$, $P = 0.202$), indicating that there was no measurable difference in performance for upright and inverted targets. We did however find a significant main effect of crowding condition ($F[2,12] = 127.35$, $P < 0.0001$), as well as a significant interaction between the two factors ($F[2,12] = 4.32$, $P = 0.039$), which we investigated further with planned Bonferroni-corrected paired t-tests. For an upright target face, the identification of horizontal eye-shifts was strongly impaired with upright flankers and significantly less impaired with inverted ($t[6] = -3.32$, $P = 0.032$), as with identity judgements in Experiment 1. Crucially, performance with an inverted target was strongly impaired by inverted flankers and significantly less impaired by upright flankers ($t[6] = -3.20$, $P = 0.038$), unlike the pattern observed for inverted targets in Experiment 1.

We find that crowding is modulated by the orientation of the flankers for both upright and inverted target faces. This is contrary to the prediction of the holistic account of crowding that there should be no effect of flanker orientation for stimuli that are not processed holistically, such as inverted faces¹⁹. Furthermore, our task did not produce a significant difference between upright and inverted targets when uncrowded. If this lack of an inversion effect was taken to indicate a lack of holistic engagement, then according to Louie *et al.*¹⁹ we should have found no flanker modulation of crowding even with upright target faces. In fact we find the opposite: both upright and inverted target faces produced a pattern of results resembling that seen for upright faces in the identity-matching task of Experiment 1.

We argue that this symmetric pattern of target-flanker similarity emerged here due to the matched task difficulty between upright and inverted target conditions (which can be seen with uncrowded performance). This is in line with our account of crowding being selective for target-flanker similarity in all stimuli, including faces.

Alternatively, it could be argued that this general selectivity emerged in Experiment 2 because of the switch in task from identity judgements to eye judgements. Indeed it has been argued that the nature of the task can be central to the pattern of crowding²⁶. As such, the selectivity for target-flanker similarity observed in this experiment could be due to the eye-judgement task primarily engaging featural (rather than holistic) processes. To rule out this interpretation, we next examined crowding using a vertical eye-judgement task that is susceptible to the inversion effect.

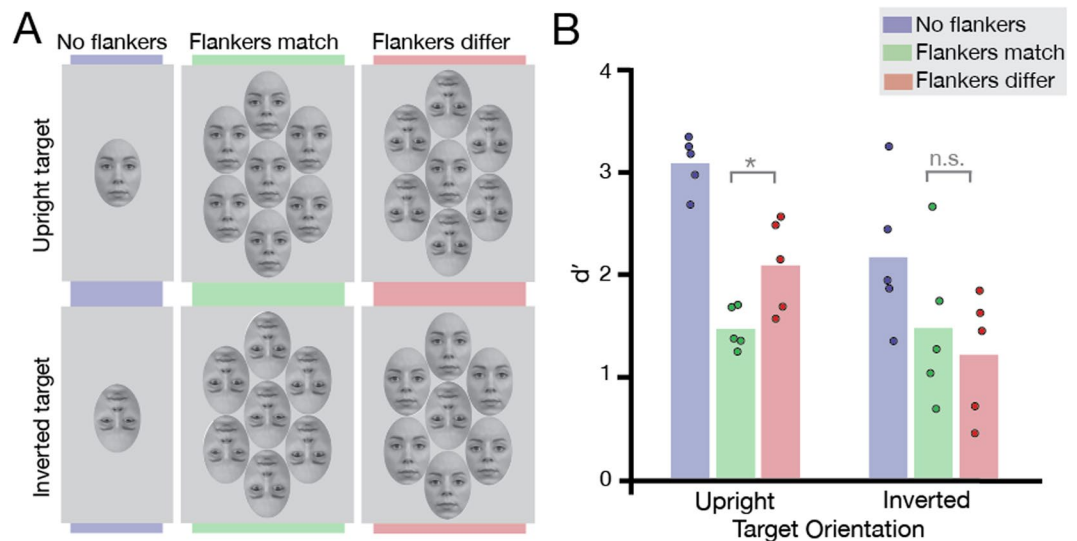


Figure 3. Design and results of Experiment 3 with vertical eye-judgements. **(A)** *Experimental conditions:* Two target orientations were presented: upright (top panel) and inverted (bottom panel). The crowding conditions were identical to Experiments 1 & 2. The target face could have the same vertical eye positions with the reference (“normal”, no eye-shift), or different (eyes shifted down). Face stimuli were taken from the Radboud Faces Database³² and edited. **(B)** *Results of Experiment 3:* Results plotted for the three crowding conditions, for upright and inverted targets ($N = 5$). Bars depict average d' values; dots show d' for each observer. Grey brackets indicate the significance level of each condition comparison (n.s. not significant, $*P < 0.05$).

Experiment 3

Although horizontal eye-shifts are minimally affected by inversion, vertical eye shifts have been shown to be disrupted in a similar fashion to identity judgements²⁸. If it is the increased task difficulty resulting from inversion that obscures selectivity for target-flanker similarity, then we should find a pattern of results comparable to that observed with identity judgements (as seen in Experiment 1 and Louie *et al.*¹⁹). Specifically, the susceptibility of vertical-eye shifts to the inversion effect should give a modulation of crowding by flanker orientation for upright target faces, but not for inverted target faces. In contrast, if the change from identity judgements to more featural eye-shift judgements is responsible for the symmetric pattern of crowding in Experiment 2, then the same symmetric pattern should be observed with vertical eye-judgements: crowding should be modulated by the orientation of the flankers for both upright and inverted faces.

Methods. All experimental details were identical to Experiment 2, with the following exceptions. Five observers (2 males, $M_{\text{age}} = 26$) took part, all with normal or corrected-to-normal vision. To construct the stimuli for this experiment we shifted the eyes and eyebrows of the reference face (RaDF-090-12³²) by 20 pixels along the vertical axis. This face served as the target in the trials in which the target differed from the reference face (with unaltered eye positions), as well as three of the six flanker faces in crowded trials (see Fig. 3A). Stimulus sizes varied between observers, ranging from 2.3–4.0° ($M = 3^\circ$) horizontally, and 3.3–5.8° ($M = 4.6^\circ$) vertically. This resulted in on-screen vertical eye-shifts for each observer ranging from 0.17–0.24° ($M = 0.21^\circ$). As in Experiment 2, we tested observers’ performance on two target orientations and three crowding conditions (see Fig. 3A).

Results and Discussion. We again calculated d' as a measure of sensitivity to differences in vertical eye-position (see Fig. 3B). These values were analysed using a two-way repeated measures ANOVA. Analyses revealed a main effect of crowding condition ($F[2, 8] = 38.10, P < 0.0001$). There was no main effect of target orientation ($F[1, 4] = 3.40, P = 0.139$), despite four out of five subjects performing worse with inverted target faces than upright, similar to prior work²⁸. We suspect this is due to the large variability in performance for inverted targets, rather than the lack of an inversion effect *per se*, and indeed the large effect size ($\eta^2_{\text{partial}} = 0.46$) suggests that this may be the case. Additionally, the ANOVA also revealed a significant interaction between target orientation and crowding condition ($F[2, 8] = 6.44, P = 0.022$). Planned paired t-tests with Bonferroni corrections showed that crowding followed the same pattern as in identity-matching tasks: crowding was strong for upright targets with upright flankers and significantly weaker with inverted flankers ($t[4] = -3.58, P = 0.046$). As in Experiment 1, there was no modulation of crowding by flanker orientation for inverted targets, with similarly poor performance for both upright and inverted flankers ($t[4] = 0.88, P = 0.86$).

We thus replicate the “holistic similarity” pattern of crowding with an eye-judgement task, as previously observed for identity judgements in both the study by Louie *et al.*¹⁹ and in our replication in Experiment 1. As such, eye-judgement tasks can also produce the pattern of results previously attributed to holistic crowding processes. Consequently, a switch in task²⁶, from identity- to eye-judgements cannot explain the differing pattern of selectivity that arose in Experiment 2. We argue that the common factor in these experiments is the role of task difficulty – crowding is modulated by the orientation of the flankers when uncrowded target identification

is easy (for upright target faces in all tasks and for horizontal-eye judgements with inverted targets), and is obscured when the task is difficult (for identity and vertical-eye judgements with inverted targets). Were this the case, it should be possible to obscure this modulation of crowding even for an upright target face, simply by increasing task difficulty. In other words, with increased difficulty in a task where crowding shows selectivity for target-flanker similarity, it should be possible to “turn off” the modulatory effects of the flankers. In Experiment 4 we test this by increasing the difficulty of horizontal eye-judgements, a manipulation that also allows us to further explore the role of holistic processes in eye-judgement tasks.

Experiment 4

Here, our primary aim was to investigate the role of task difficulty on judgements of horizontal eye position within upright target faces. Varying the amount of horizontal displacement allows a straightforward manipulation of task difficulty, without changing parameters such as eccentricity or target-flanker separation that would affect the magnitude of crowding⁵. If task difficulty can obscure the selectivity of crowding for target-flanker similarity in inverted target faces (making both upright and inverted flankers equally disruptive to identification), then by increasing the difficulty of our horizontal eye-judgement task (by decreasing the displacement of the eyes), it should also be possible to disrupt the selectivity for target-flanker similarity in *upright* target faces. Were this possible, upright and inverted flankers should produce equivalent levels of crowding for an upright target. Alternatively, if task difficulty does not modulate the effect of crowding on faces, then the recognition of upright target faces should be more disrupted by upright than inverted flankers in all cases where uncrowded performance is above chance. We examine this possibility in Experiment 4a (Fig. 4A).

Previous studies³⁷ have in fact demonstrated that when differences in facial features are more subtle, holistic processes are more strongly engaged. As such, horizontal eye judgements with a reduced displacement in the eyes should not only become more difficult, but may also engage holistic processes to a *greater* degree than the larger displacements in Experiment 2. If crowding is modulated by holistic similarity, this increased holistic engagement should increase (or at least maintain) the difference in crowding between upright and inverted flanker conditions. In order to assess this possibility, in Experiment 4b (Fig. 4C) we compared uncrowded performance with upright and inverted targets for both small and large eye displacements to seek the presence of an inversion effect (the classic indicator of holistic engagement^{20–24}). If horizontal eye judgements can engage holistic processes when feature differences are subtle, we should find an inversion effect with small eye displacements. If crowding is sensitive to holistic similarity, the presence of an inversion effect with these small displacements should coincide with a difference between the upright and inverted flanker conditions in Experiment 4a. In contrast, an inversion effect for small displacements in Experiment 4b in the absence of any difference between flanker conditions in Experiment 4a would provide further evidence that crowding does not rely on holistic similarity.

Methods. *Experiment 4a.* Methods were identical to those of Experiment 2, with the following exceptions. Five observers (2 females, $M_{\text{age}} = 27.4$ years) who participated in Experiment 2 took part. To increase task difficulty, the displacement of the horizontal eye position was reduced compared to Experiment 2: the eyes and eyebrows of face RaDF-090-12³² were shifted inwards by 10 pixels (half the displacement of Experiment 2). Stimulus sizes varied between 2.1–3.6° ($M = 3.1^\circ$) horizontally and 3.0–5.2° ($M = 4.5^\circ$) vertically, leading to on-screen horizontal eye displacements of 0.08–0.13° ($M = 0.11^\circ$). As in Experiment 2, observers were shown peripheral target faces at 12° eccentricity that were either unaltered or shifted inwards, and were required to judge whether the target face was the same or different from the reference with regards to horizontal eye position. As the target face was always upright, this resulted in three crowding conditions: “uncrowded”, “crowded-same, and “crowded-different” and 720 trials (3 crowding conditions \times 2 same-different conditions \times 30 repetitions per face \times 4 blocks).

Experiment 4b. Five observers (2 males, $M_{\text{age}} = 31.5$ years) took part, all with normal or corrected-to-normal vision. Stimuli were as in Experiment 4a, though only the target-alone uncrowded conditions were used. Stimulus sizes varied between 2.0–3.9° ($M = 3.1^\circ$) horizontally and 2.9–5.4° ($M = 4.5^\circ$) vertically. This led to on-screen horizontal eye displacements of 0.15–0.27° ($M = 0.22^\circ$) for the “large shift” condition and 0.07–0.14° ($M = 0.11^\circ$) for the ‘small shift’ condition. As in Experiments 2 and 4a, observers judged whether the target, presented at 12° eccentricity, was the same or different from the reference face in terms of eye position. The target was always presented in isolation (“uncrowded”) and was either upright or inverted, resulting in 480 trials per eye shift condition (2 same-different conditions \times 30 repetitions per face \times 4 blocks).

Results and Discussion. As in Experiment 2, d' was calculated as a measure of observers’ sensitivity to differences in eye separation. Results for Experiment 4a are shown in Fig. 4B. It is clear that the smaller eye shifts were harder to detect even when the target was uncrowded (plotted in Fig. 4B against data with large displacements from Experiment 2). Nonetheless, crowding still had a significant effect on identifying differences in horizontal eye displacement: Bonferroni corrected t-tests showed that performance with uncrowded targets was better than in conditions where the target was surrounded by flankers of the same ($t[4] = 8.23, P = 0.003$) or different orientation ($t[4] = 6.04, P = 0.012$). Importantly however, when crowded, there was no significant difference between observers’ ability to detect small eye displacements with upright vs. inverted flankers ($t[4] = 0.5, P > 0.99$). The equivalent difficulty of these two conditions resembles the pattern of results for inverted targets in identity-matching tasks (Experiment 1).

These results provide direct evidence that an increase in task difficulty can obscure the selectivity of crowding for target-flanker similarity, even for upright target faces. Here we produce this effect using the same horizontal eye-judgements previously shown in Experiment 2 to produce a clear selectivity for target-flanker similarity. We argue that as performance drops for uncrowded targets, the range over which performance can vary in crowded

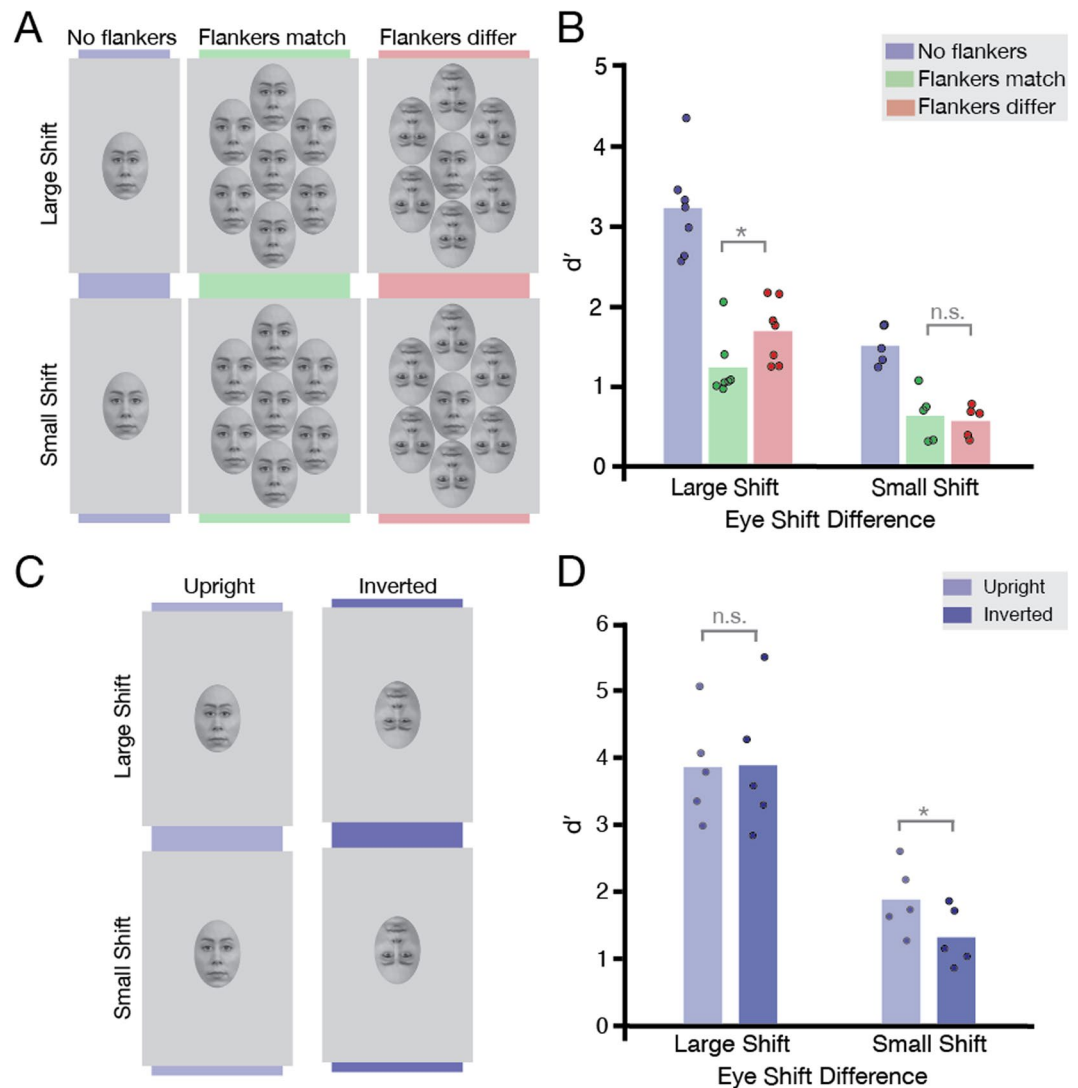


Figure 4. Design and results of Experiments 4a and 4b, designed to test the role of task difficulty. **(A)** *Experimental conditions of Experiment 4a:* Only upright targets were used, tested in the same crowding conditions as before: either in isolation or surrounded by either upright or inverted flankers. The top panel shows the large eye-shifts used in Experiment 2; the bottom panel shows the small eye-shifts used in Experiment 4a. The horizontal eye positions of the target were either the same (“normal”, no eye-shift) or different (inward eye shift) to the reference. Face stimuli were taken from the Radboud Faces Database³² and edited. **(B)** *Results of Experiment 4a:* Results plotted for upright targets in the three crowding conditions, replotted with large eye-shifts from Experiment 2 ($N = 7$) and small shifts from the current experiment ($N = 5$). Bars indicate average d' ; dots indicate d' for each observer. Grey brackets indicate the significance level of each condition comparison (n.s. not significant, $*P < 0.05$). **(C)** *Experimental conditions of Experiment 4b:* Targets were always presented in isolation, either upright or inverted. Upper panels show the large horizontal eye-shifts; lower panels show the small ones. **(D)** *Results of Experiment 4b:* Results plotted for upright and inverted targets for the two eye-shift conditions ($N = 5$). Bars indicate average d' ; dots indicate d' for each observer. Grey brackets indicate the significance level of each condition comparison (n.s. not significant, $*P < 0.05$).

conditions becomes restricted, thereby obscuring any differences between the upright and inverted flanker conditions. We explore this mechanism further in the General Discussion.

One criticism is that horizontal eye-judgement tasks may not engage holistic processes, and that the pattern observed in Experiments 2 and 4a is a result of this. The results of Experiment 4b, plotted in Fig. 4D, contradict this view. As in Experiment 2, the identification of large eye-shifts is equally difficult in upright and inverted faces, with no significant difference in performance between upright and inverted target conditions ($t[4] = -0.14$, $P = 0.849$, Bonferroni corrected). However, with small eye-shifts, we find a significant difference in performance between upright and inverted target conditions ($t[4] = 3.76$, $P = 0.039$, Bonferroni corrected), with small eye shifts being more difficult to identify in inverted faces. This inversion effect indicates that holistic processes are engaged when horizontal eye judgements are sufficiently subtle, in line with previous findings demonstrating a greater engagement of holistic processes when the discriminability of facial features is decreased³⁷.

We note that overall performance in the small eye-shift condition is lower than that reported in previous studies using eye shifts of similar magnitude, for stimuli presented foveally^{28,38}. This is likely due to the peripheral presentation of our stimuli, with resulting decreases in visibility via reduced acuity and contrast sensitivity², and a probable increase in within-face crowding³⁵. Our inversion effect is similarly smaller than that found in these prior studies, though it is consistent with the range of values found more generally^{39,40}. Nonetheless, the inversion effect that we observe, along with those found in prior studies^{28,38}, suggests the potential for holistic processes to be engaged in judgements of horizontal eye separation.

These findings provide further evidence that the differing pattern of results in Experiments 1 and 2 did not arise simply due to our use of an identity task in the former and an eye-judgement task in the latter. In Experiment 4b, with small eye-shifts, our horizontal eye-judgement task shows a significant inversion effect, the typical marker of holistic engagement^{20–24}. If the “holistic similarity” criterion of crowding were to operate in a similar fashion, then we should have observed a *stronger* modulation of crowding in the small eye-shift condition of Experiment 4a – that is, the difference between upright and inverted flanker conditions should have increased (or at the very least, to have remained present). In contrast, we find no modulation of crowding by the orientation of the flankers on an upright face. This finding cannot be attributed to the task engaging primarily local processes, given the presence of a significant inversion effect for the same task in Experiment 4b. Rather, our findings together demonstrate that the common factor that obscures the release from crowding is task difficulty, rather than any propensity for crowding to operate holistically – with high difficulty there is a lack of flanker modulation for crowding (here for upright faces), which can be revealed when difficulty is lowered (as in Experiment 2, with inverted faces).

Experiment 5

If task difficulty can account for the asymmetry in the crowding of faces (seen in Experiments 1 and 3), and holistic similarity does not determine the strength of crowding, what is driving the difference in crowding with upright and inverted flankers when it occurs? It is important to consider here that inversion not only reduces the capacity for the holistic processing of faces, but also alters the features within faces. Relative to an upright face, inversion alters the orientation of each feature (i.e. the nose rotates to be upside-down), as well as the spatial order of these features (the eyes above nose above mouth pattern is reversed)^{28,41}. More specifically, the spatial order of the features results in a predominantly top-heavy pattern in upright faces that becomes predominantly bottom-heavy (in a retinotopic sense) when they are inverted^{42,43}. Similar differences in feature positions have previously been shown to modulate crowding within letter-like stimuli – an upright T will be crowded less by inverted T flankers than by other configurations⁴⁴, for instance. In our final experiment, we manipulated both the orientation of facial features and their spatial order independently in order to assess the contribution of these dimensions to the selectivity of face crowding for target-flanker similarity.

Methods. Methods were similar to Experiment 2, with the following exceptions. Five observers (4 females, $M_{\text{age}} = 25$) took part, all with normal or corrected-to-normal vision. The target and reference faces were always the face RaDF-090-12³², either with no eye shift or with the eyes and eyebrows displaced by 20 pixels inwards (as in Experiment 2). Stimulus sizes for each observer ranged between 2–4° ($M = 3.2^\circ$) horizontally and 2.9–5.8° ($M = 4.7^\circ$) vertically. This resulted in horizontal eye displacements between 0.15–0.36° ($M = 0.26^\circ$) for stimuli presented on-screen.

We used the reference face with and without the horizontal displacements to manipulate the features and first-order relational properties⁴¹ of the flankers independently. To examine the effect of feature orientations, we rotated the facial features of the faces (to match an inverted face), whilst keeping their positions in the same spatial order as the upright target (Fig. 5A, bottom panel). To examine the effect of the spatial order of facial features, we shifted their position in the faces (to match inverted faces) without rotating them (to match the upright target; Fig. 5A, top right panel). These two new “Thatcherised”⁴⁵ faces, as well as the upright and inverted faces used in Experiment 2, served as the four different flanker conditions in this experiment.

In this experiment, the target face was always upright, with five crowding conditions, including “uncrowded”. These conditions with their relevant face flankers can be seen in Fig. 5A. In the first two crowded conditions, flankers were identical to the previous experiments: either upright flankers surrounding an upright target, or inverted flankers surrounding an upright target. In the third “inverted features” condition, the feature orientations were altered (i.e. rotated to be consistent with an inverted face) but their positions were held constant (consistent with the spatial order of features in an upright face). For the final “inverted positions” condition, the facial feature orientations were held constant (to be consistent with an upright face) while the positions were rotated (to match the spatial order of an inverted face). This resulted in 1200 trials (5 crowding conditions \times 2 same/different conditions \times 15 repetitions per face \times 2 target faces \times 4 blocks).

Results and Discussion. Observers’ performance in detecting horizontal eye-shifts was measured in the five flanker conditions (Fig. 5A), with d' again computed as an indicator of sensitivity to differences in eye-separation (Fig. 5B). A one-way repeated measures ANOVA revealed a significant main effect of crowding condition ($F[4,16] = 99.85$, $P < 0.0001$). Planned t-tests with Bonferroni corrections revealed that as in the other experiments, there was strong crowding when an upright-face target was surrounded by upright flankers and significantly weaker crowding with inverted flankers ($t[4] = -6.001$, $P = 0.012$). Importantly, when flankers had rotated features but kept the spatial order of an upright face (“Inverted features” in Fig. 5B), crowding remained strong and performance did not differ from that with upright flankers ($t[4] = -1.566$, $P = 0.579$). Crowding was however weakened (leading to improved performance) when flankers had upright features re-arranged to match the spatial order of an inverted face (“Inverted positions” in Fig. 5B) – performance in this condition was significantly better than performance with upright flankers ($t[4] = -8.54$, $P = 0.003$).

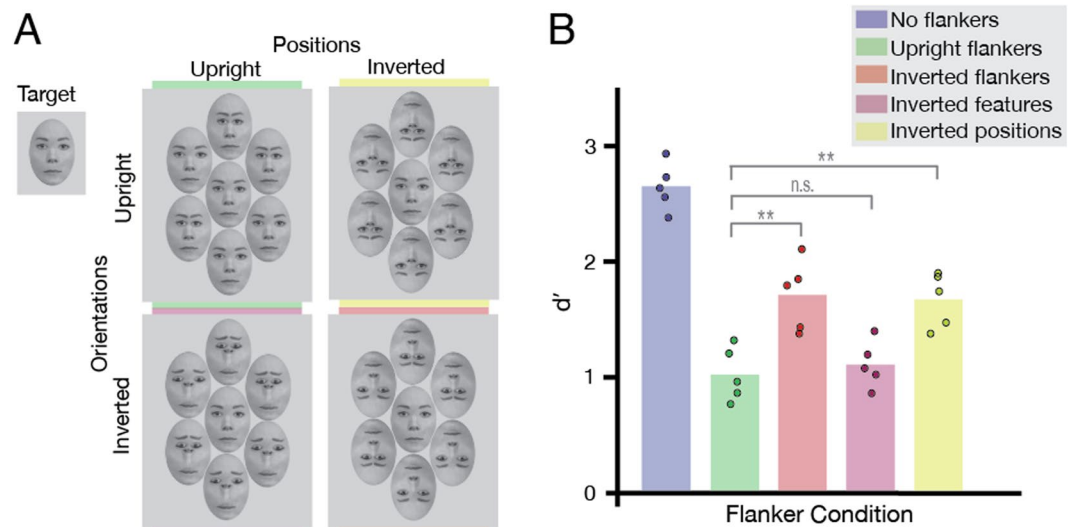


Figure 5. Design and results of Experiment 5, designed to test the selectivity of face crowding. **(A)** *Experimental conditions:* Only upright target faces were presented, with five flanker conditions: no flankers, flankers with upright positions and features (top left, upright flankers), flankers with inverted positions and features (bottom right, inverted flankers), flankers with inverted features but upright positions (bottom left, “inverted features”), and flankers with upright features but inverted positions (top right, “inverted positions”). Face stimuli were taken from the Radboud Faces Database³² and edited. **(B)** *Results of Experiment 5:* Results plotted for the five flanker conditions for upright targets (N = 5). Bars depict average d' ; dots indicate d' for each observer. Grey brackets indicate the significance level of each condition comparison (n.s. not significant, * $P < 0.05$, ** $P < 0.01$).

Our findings thus demonstrate that it is the differences in the spatial order of facial features between target and flanker elements that drives the selectivity of upright face crowding. Although this spatial order is likely a key aspect for the engagement of holistic processing^{21–23}, the results of Experiments 2, 4a, and 4b indicate that these differences in the spatial order between target and flanker elements can modulate crowding regardless of the degree of holistic engagement by the task. If this task were to exclusively drive featural (rather than holistic) crowding³⁵, we should have observed a release from crowding with the inversion of facial features in the flankers. However, this condition produced an equivalent degree of crowding to upright flanker faces. The differences in crowding that we observe with inverted flanker faces and flankers with inverted feature positions (but upright feature orientations) suggest that it is the spatial order of faces that modulates crowding between faces, not the extent of holistic engagement.

Because upright faces are top-heavy stimuli, inversion changes the distribution of this content across the image, making inverted faces bottom-heavy^{42,43}. When the spatial order of flanker features is changed by inverting their positions (whilst keeping their orientation constant), the flanker faces become similarly bottom heavy, again leading to a release from crowding with an upright (top-heavy) face. In contrast, changing the orientation of the features whilst maintaining upright spatial order in the flankers maintains the top-heavy content, and crowding remains strong. We have further quantified these differences in image content (see Supplementary Materials and Supplementary Fig. 1), and show that spatial order variations can well account for our results in this experiment. We further demonstrate that changes in the top-heavy nature of faces can also account for findings regarding “holistic crowding” obtained with Mooney faces²⁵ (Supplementary Fig. 2). The lack of orientation selectivity observed by Louie *et al.*¹⁹ for houses may conversely reflect the more vertically balanced nature of the image content in this stimulus set – these cropped sections of house images are unlikely to produce a consistent difference in image content between upright and inverted elements. In sum, we argue that it is differences in the spatial order within target and flanker stimuli that modulates crowding, regardless of their propensity for holistic processing.

General Discussion

We have shown that crowding is driven by a general selectivity for target-flanker similarity, even in complex stimuli such as faces. In Experiment 1, we replicated the asymmetric pattern of crowding observed in Louie *et al.*¹⁹ – for these identity judgements, crowding was strong for an upright target face surrounded by upright flankers, and weak with inverted flankers. In contrast, with an inverted target face, crowding was strong with both upright and inverted flankers. This asymmetry was previously argued to reflect the selectivity of an additional stage of crowding sensitive to holistic similarity in faces^{7,19,25}, independent from the crowding of inverted faces and simple objects.

On the contrary, in Experiment 2, we examined judgements of horizontal eye separation, known to show minimal inversion effects²⁸, and found that crowding was strong when the orientation of the flankers matched the target and weak when it differed for *both* upright and inverted target faces. The symmetric selectivity that we observe with this task is opposed to the predictions of the “holistic similarity” account, whereby crowding

should be modulated by the orientation of the flankers only for an upright holistically-processed target face. Our observation of this pattern with both upright and inverted target faces is therefore inconsistent with a distinct process for the crowding of upright faces. Rather, our results follow the more general pattern seen for a range of fundamental visual features, including orientation^{6,9,10}, colour^{11–13}, position¹⁸, motion¹³, and spatial frequency¹⁷, suggesting the operation of a general mechanism.

Why then does crowding show an asymmetric pattern of selectivity in some tasks but not others? Experiment 4a shows that when horizontal eye shifts become more difficult to detect, crowding for an upright target face is as strong with inverted flankers as with upright flankers, a pattern that ordinarily occurs for inverted faces in tasks such as identity matching¹⁹ (replicated in Experiment 1) and vertical eye-judgements (Experiment 3). Experiment 4b further demonstrates that these small horizontal eye shifts actually engage holistic processes to a greater extent than do larger shifts (consistent with prior work³⁷). If crowding were operating holistically, one would therefore expect greater modulation of crowding by the orientation of the flankers in this case, rather than the lack of modulation that we observe. Rather, it is the increase in task difficulty, and not a lack of holistic processing, that can obscure the symmetric selectivity of crowding for target-flanker similarity, even for upright faces. The lack of apparent selectivity in these cases of high task difficulty can be attributed to a restriction in the range over which observers' performance varies. By decreasing performance for uncrowded targets, task difficulty restricts the upper bound for the release from crowding. In the case of an inverted target face, the difficulty in making identity judgements obscures these modulations because the decline in uncrowded performance leaves little room for improvement.

In fact, similar effects of task difficulty can also be observed in previous studies using judgements of identity and gender that clearly engage holistic processes. In Experiment 1 of Louie *et al.*¹⁹, increasing levels of random dot noise were applied to the upright target face, making identity judgements gradually more difficult. With this increase in difficulty, crowding became equally strong with upright and inverted flankers at the fourth level of noise used in their study (their Fig. 1D). This effect was not due to a floor effect with uncrowded targets, as uncrowded performance remains above chance levels, meaning that the addition of the flankers could still make identity judgements harder. Nonetheless, upright target faces in this case were no longer modulated by the orientation of the flankers. A similar effect can be observed for gender judgements used to examine crowding with Mooney faces²⁵. Farzin *et al.*²⁵ found that the recognition of uncrowded upright Mooney faces became more difficult with increasing eccentricity (their Figure 8C). Although an effect of target-flanker differences in face orientation was observed at eccentricities of 3° and 6°, at 10° these judgements became sufficiently difficult that the modulation disappeared. Once again, uncrowded judgements remain above chance, and yet the pattern of selectivity for target-flanker similarity disappears.

We have explored the effect of task difficulty further with a population model that is similar in principle to recent models of the effect of crowding on orientation^{46,47}. By simulating a population response to interocular eye distance, consistent with psychophysical^{48,49} and physiological findings^{50,51}, we can similarly model the crowding of faces as a pooling process that combines responses to the target and flanker elements. With this model, we demonstrate that an increase in task difficulty (produced by decreasing the interocular separation for horizontal eye judgements, as in Experiment 4) can decrease the separability of target and flanker signals (see Supplementary Fig. 3). Although an increase in target-flanker differences would ordinarily allow the target response to be clearly separated from flanker responses, an increase in task difficulty must be accompanied by a greater reduction in crowding to give the same separation in the population response. Similar complications would likely arise for judgements of identity with inverted faces. More generally, these effects of task difficulty may also explain “supercrowding” effects, where the masking of a target can impair its identification beyond the typical range of interference zones for crowding⁵².

Our final experiment demonstrates that the crowding of faces is driven by a selectivity for the spatial order of facial features. Crowding is strong when the spatial order of facial features (i.e. their first-order relations) are matched between target and flankers, and weak when they differ, regardless of individual feature orientations. This effect of spatial structure is broadly consistent with findings⁵³ that schematic faces and electrical-plug flankers crowd photographic faces more than Chinese characters (although these stimuli did not produce an effect of flanker orientation, likely due to the differences in texture, contrast, and spatial frequency content between target and flanker elements). The spatial order of facial features is a key aspect of holistic processes in face perception^{21–23}, perhaps driven by the top-heavy^{42,43} vertical ordering of horizontal structure^{54–56}. However, the results of Experiment 2 demonstrate that crowding is selective for target-flanker differences in this spatial order even with inverted target faces, irrespective of their capacity for holistic processing. With the photographic face stimuli used in our study, this selectivity could be driven by image-based differences in the vertical configuration of horizontally oriented image structure (Figs S1 and S2). Although these differences in spatial structure clearly drive higher-level areas⁵⁷, spatial variations of this nature could be processed in cortical areas as early as V1⁵⁸, V2^{59,60} or V4⁶¹, all of which have been implicated in the general operation of crowding^{62–67}. Spatial order has also been shown as relevant for crowding more generally, through its modulation by contours^{68,69}, spatial regularity^{70–72}, and the structure of letter-like stimuli⁴⁴. There is therefore no need to invoke multiple processes of crowding to explain the selectivity for target-flanker similarity in upright faces – changes in the spatial order of features within upright or inverted faces can modulate the strength of crowding via these more general processes.

Our argument that face crowding can be explained by a combination of selectivity for spatial order and task difficulty could also explain similar “holistic” findings for crowding. In particular, crowding for upright Mooney faces⁷³ has been shown to be strong with upright Mooney flankers, and weak with inverted flankers²⁵. As with photographic faces (Fig. S1), our image analyses show that upright Mooney faces are top-heavy stimuli, with the majority of their contours concentrated at the top of the face image, whereas inverted Mooney faces are bottom-heavy stimuli (Fig. S2). Furthermore, although the visibility of individual features is degraded in Mooney faces, prior work shows that their contours remain susceptible to within-face crowding in much the same fashion

as photographic face images (see Experiment 3 of Farzin *et al.*²⁵). This identical susceptibility to inter-feature crowding suggests that similar processes can account for crowding in both stimulus types. A shift from top- to bottom-heavy spatial order in the Mooney flankers (relative to an upright target) would therefore be expected to reduce crowding, similar to the spatial-order effects seen with other stimuli^{44,68–71,72}. As such, target-flanker differences in spatial order, rather than their propensity for holistic processing, could also drive the crowding of Mooney faces. The shift in the spatial order of image content could also drive crowding for other stimuli that have canonical upright configurations like faces, such as Chinese characters⁷⁴ and biological-motion arrays⁷⁵.

Altogether, our results demonstrate that when task difficulty is matched between upright and inverted target conditions, crowding for faces is selective for differences in the spatial order of facial features. We argue that this reflects a general selectivity for target-flanker similarity that determines the strength of interactions between stimuli ranging from faces¹⁹ to facial features (“within-face” crowding)³⁵ and simple objects^{8,16,18}. In other words, there is no need for an independent high-level crowding process that is selective for holistic similarity^{19,25} – crowding is likely to operate in the same way for all stimuli. Note that this does not contradict evidence for other higher-level effects in crowding, including the Gestalt grouping processes that have been shown to modulate crowding in multi-element arrays^{70,76,77}. Additionally, our account is not necessarily inconsistent with the involvement of multiple visual processing stages^{7,26} or multiple visual brain areas^{62,63,66,78} in crowding. Given the wide range of stimuli susceptible to crowding (from colour patches and moving Gabors to faces), distinct cortical regions or populations of neurons may be involved in the computations that produce the disruption within each feature domain. In this sense, we cannot entirely exclude the possibility that the crowding of faces occurs in a distinct stage – we simply demonstrate here that the behavioural evidence for this possibility (based on differences in the selectivity of crowding) does not hold up to scrutiny. Importantly, even if separate processing stages are required for crowding, our results suggest that each should operate in the same fashion – with a common selectivity for the visual similarity between target and flanker objects.

References

1. Anstis, S. M. Picturing peripheral acuity. *Perception* **27**, 817–825 (1998).
2. Rosenholtz, R. Capabilities and limitations of peripheral vision. *Annual Review of Vision Science* **2**, 437–457 (2016).
3. Flom, M. C., Weymouth, F. W. & Kahneman, D. Visual resolution and contour interaction. *Journal of the Optical Society of America* **53**, 1026–1032 (1963).
4. Bouma, H. Interaction effects in parafoveal letter recognition. *Nature* **226**, 177–178 (1970).
5. Toet, A. & Levi, D. M. The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research* **32**, 1349–1357 (1992).
6. Levi, D. M., Hariharan, S. & Klein, S. A. Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant or simple contrast masking. *Journal of Vision* **2**, 3 (2002).
7. Whitney, D. & Levi, D. M. Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences* **15**, 160–168 (2011).
8. Andriessen, J. J. & Bouma, H. Eccentric vision: Adverse interactions between line segments. *Vision Research* **16**, 71–78 (1976).
9. Hariharan, S., Levi, D. M. & Klein, S. A. “Crowding” in normal and amblyopic vision assessed with Gaussian and Gabor C’s. *Vision Research* **45**, 617–633 (2005).
10. Wilkinson, F., Wilson, H. R. & Elleberg, D. Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A* **14**, 2057–2068 (1997).
11. Kennedy, G. J. & Whitaker, D. The chromatic selectivity of visual crowding. *Journal of Vision* **10**, 15 (2010).
12. Pöder, E. Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research* **71**, 641–645 (2007).
13. Gheri, C., Morgan, M. J. & Solomon, J. A. The relationship between search efficiency and crowding. *Perception* **36**, 1779–1787 (2007).
14. van den Berg, R., Roerdink, J. B. T. M. & Cornelissen, F. W. On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision* **7**, 14 (2007).
15. Bex, P. J. & Dakin, S. C. Spatial interference among moving targets. *Vision Research* **45**, 1385–1398 (2005).
16. Kooi, F. L., Toet, A., Tripathy, S. P. & Levi, D. M. The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision* **8**, 255–279 (1994).
17. Chung, S. T., Levi, D. M. & Legge, G. E. Spatial frequency and contrast properties of crowding. *Vision Research* **41**, 1833–1850 (2001).
18. Greenwood, J. A., Bex, P. J. & Dakin, S. C. Crowding follows the binding of relative position and orientation. *Journal of Vision* **12**, 1–20 (2012).
19. Louie, E. G., Bressler, D. W. & Whitney, D. Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision* **7**, 24–24 (2007).
20. Tanaka, J. W. & Farah, M. J. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: A Human Experimental Psychology* **46**, 225–245 (1993).
21. Rossion, B. Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica* **128**, 274–289 (2008).
22. Young, A. W., Hellawell, D. & Hay, D. C. Configural information in face perception. *Acta Psychologica* **128** (1987).
23. Le Grand, R., Mondloch, C. J., Maurer, D. & Brent, H. P. Neuroperception: Early visual experience and face processing. *Nature* **410**, 890 (2001).
24. Yin, R. K. Looking at upside-down faces. *Journal of Experimental Psychology* **81**, 141–145 (1969).
25. Farzin, F., Rivera, S. M. & Whitney, D. Holistic crowding of Mooney faces. *Journal of Vision* **9**, 1–15 (2009).
26. Chaney, W., Fisher, J. & Whitney, D. The hierarchical sparse selection model of visual crowding. *Frontiers in Integrative Neuroscience* **8**, 73 (2014).
27. Manassi, M. & Whitney, D. Multi-level crowding and the paradox of object recognition in clutter. *Current Biology* **28** (2018).
28. Goffaux, V. & Rossion, B. Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception & Performance* **33**, 995–1001 (2007).
29. Brainard, D. H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433–436 (1997).
30. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437–442 (1997).
31. Kleiner, M., Brainard, D. & Pelli, D. What’s new in Psychtoolbox-3? *Perception* **36** ECVP Abstract Supplement (2007).
32. Langner, O. *et al.* Presentation and validation of the Radboud Faces. *Cognition & Emotion* **24**, 1377–1388 (2010).
33. Ellis, H. D., Shepherd, J. W. & Davies, G. M. Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception* **8**, 431–439 (1979).
34. Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M. & Ellis, A. W. Matching familiar and unfamiliar faces on internal and external features. *Perception* **14**, 737–746 (1985).

35. Martelli, M., Majaj, N. J. & Pelli, D. G. Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision* **5**, 58–70 (2005).
36. Macmillan, N. A. & Creelman, C. D. Detection theory: A user's guide. 2 edn, (Lawrence Erlbaum Associates, 2005).
37. Goffaux, V. The discriminability of local cues determines the strength of holistic face processing. *Vision Research* **64**, 17–22 (2012).
38. Crookes, K. & Hayward, W. G. Face inversion disproportionately disrupts sensitivity to vertical over horizontal changes in eye position. *Journal of Experimental Psychology: Human Perception and Performance* **38**, 1428–1437 (2012).
39. McKone, E. Isolating the special component of face recognition: Peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 181–197 (2004).
40. McKone, E. & Yovel, G. Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review* **16**, 778–797 (2009).
41. Diamond, R. & Carey, S. Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General* **115**, 107–117 (1986).
42. Viola Macchi, C., Turati, C. & Simion, F. Can a nonspecific bias towards top-heavy patterns explain newborns' face preference? *Psychological Science* **15**, 379–383 (2004).
43. Simion, F., Valenza, E., Macchi Cassia, V., Turati, C. & Umiltà, C. A. Newborn's preference for up-down asymmetrical configurations. *Developmental Science* **5**, 427–434 (2002).
44. Dakin, S. C., Cass, J., Greenwood, J. A. & Bex, P. J. Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *J Vis* **10**, 14, <https://doi.org/10.1167/10.10.14> (2010).
45. Thompson, P. Margaret Thatcher: A new illusion. *Perception* **9**, 483–484 (1980).
46. Harrison, W. J. & Bex, P. J. A unifying model of orientation crowding in peripheral vision. *Current Biology* **25**, 3213–3219 (2015).
47. van den Berg, R., Roerdink, J. B. T. M. & Cornelissen, F. W. A neurophysiologically plausible population code model for feature integration explains visual crowding. *Plos Computational Biology* **6**, e1000646 (2010).
48. Robbins, R., McKone, E. & Edwards, M. Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. *Journal of Experimental Psychology: Human Perception & Performance* **33**, 570 (2007).
49. Susilo, T., McKone, E. & Edwards, M. What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research* **50**, 300–314 (2010).
50. Chang, L. & Tsao, D. Y. The code for facial identity in the primate brain. *Cell* **169**, 1013–1028 (2017).
51. Freiwald, W. A., Tsao, D. Y. & Livingstone, M. S. A face feature space in the macaque temporal lobe. *Nature Neuroscience* **12**, 1187–1196 (2009).
52. Vickery, T. J., Shim, W. M., Chakravarthi, R., Jiang, Y. V. & Luedeman, R. Supercrowding: Weakly masking a target expands the range of crowding. *Journal of Vision* **9**, 12 (2009).
53. Sun, H.-M. & Balas, B. Face features and face configurations both contribute to visual crowding. *Attention, Perception, & Psychophysics* **77**, 508–519 (2015).
54. Goffaux, V. & Dakin, S. C. Horizontal information drives the behavioural signatures of face processing. *Frontiers in Psychology* **1**, 143 (2010).
55. Dakin, S. C. & Watt, R. J. Biological “bar codes” in human faces. *Journal of Vision* **9**, 2 (2009).
56. Goffaux, V. & Greenwood, J. A. The orientation selectivity of face identification. *Scientific Reports* **6** (2016).
57. Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C. & Goebel, R. Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia* **81**, 1–11 (2016).
58. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology* **195**, 215–243 (1968).
59. Hegdé, J. & Van Essen, D. C. Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience* **20**, RC61–66 (2000).
60. Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience* **16**, 974–981 (2013).
61. Gallant, J. L., Braun, J. & Van Essen, D. C. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259**, 100–103 (1993).
62. Anderson, E. J., Dakin, S. C., Schwarzkopf, S. D., Rees, G. & Greenwood, J. A. The neural correlates of crowding-induced changes in appearance. *Current Biology* **22**, 1199–1206 (2012).
63. Chicherov, V., Plomp, G. & Herzog, M. H. Neural correlates of visual crowding. *Neuroimage* **93**, 23–31 (2014).
64. Chung, S. T., Li, R. W. & Levi, D. M. Crowding between first- and second- order letter stimuli in normal foveal and peripheral vision. *Journal of Vision* **7**, 10 (2007).
65. Dakin, S. C., Greenwood, J. A., Carlson, T. A. & Bex, J. P. Crowding is tuned for perceived (not physical) location. *Journal of Vision* **11**, 2 (2011).
66. Freeman, J., Donner, T. H. & Heeger, D. J. Inter-area correlations in the ventral visual pathway reflect feature integration. *Journal of Vision* **11**, 15 (2011).
67. Liu, T., Jiang, Y., Sun, X. & He, S. Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Current Biology* **19**, 127–132 (2009).
68. Livne, T. & Sagi, D. Configuration influence on crowding. *Journal of Vision* **7**, 4 (2007).
69. Glen, J. C. & Dakin, S. C. Orientation-crowding within contours. *Journal of Vision* **13**, 14 (2013).
70. Manassi, M., Sayim, B. & Herzog, M. H. Grouping, pooling and when bigger is better in visual crowding. *Journal of Vision* **12**, 13 (2012).
71. Saarela, T. P., Westheimer, G. & Herzog, M. H. The effect of spacing regularity on visual crowding. *Journal of Vision* **10**, 17 (2010).
72. Herzog, M. H., Sayim, B., Chicherov, V. & Manassi, M. Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision* **15**, 1–18 (2015).
73. Mooney, C. Age in the development of closure ability in children. *Canadian Journal of Psychology* **11**, 219–226 (1957).
74. Wong, A. C. N. *et al.* Holistic processing as a hallmark of perceptual expertise for nonface categories including Chinese characters. *Journal of Vision* **12**, 7 (2012).
75. Ikeda, H., Watanabe, K. & Cavanagh, P. Crowding of biological motion stimuli. *Journal of Vision* **13**, 20 (2013).
76. Saarela, T. & Herzog, M. H. Crowding in multi-element arrays: Regularity of spacing. *Journal of Vision* **9**, 1017 (2009).
77. Francis, G., Manassi, M. & Herzog, M. H. Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychological Review* **124**, 483–504 (2017).
78. Bi, T., Cai, P., Zhou, T. & Fang, E. The effect of crowding on orientation-selective adaptation in human early visual cortex. *Journal of Vision* **9**, 13–13 (2009).

Acknowledgements

This work is supported by grants from the UK Medical Research Council (MR/K024817/1) and the Moorfields Eye Charity (ST 14 11F). V.G. is a Research Associate with the FRS-FNRS (Belgian National Scientific Research Fund).

Author Contributions

All three authors contributed to the design of the experiments. A.K.S. conducted the experiments and analyses. All three authors were involved in conducting the image analyses, designing and programming the computational model, and writing the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30900-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Supplementary information for:

Crowding for faces is determined by visual (not holistic) similarity: Evidence from judgements of eye position

Alexandra V. Kalpadakis-Smith¹, Valerie Goffaux^{2,3,4}, & John A. Greenwood¹

¹ *Experimental Psychology, University College London, London, United Kingdom*

² *Research Institute for Psychological Science, Université Catholique de Louvain, Louvain-la-Neuve, Belgium*

³ *Institute of Neuroscience, Université Catholique de Louvain, Louvain-la-Neuve, Belgium*

⁴ *Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands*

Quantifying target-flanker similarity for our stimuli

In this section we consider the specific visual dimensions that underlie the selectivity of face crowding for target-flanker similarity. As well as reducing the propensity for holistic processing¹⁻⁴, the inversion of a face alters both the orientation of the facial features and the spatial order of these features (i.e. the eyes above nose above mouth pattern is reversed)⁵. Differences between the target and flanker faces in either of these properties could be driving the variations in crowding that we observe. In Experiment 5, we independently assessed the contribution of the orientation and order of facial features in crowding by introducing two additional types of “Thatcherised”⁶ flanker faces. First, to examine the role of feature orientation, we constructed flankers with rotated facial features (to match an inverted face) where the position of features was maintained in the same first-order pattern as an upright face (“inverted features”). Second, to examine the role of the spatial order of the features, we shifted the position of facial features (to match inverted faces) without rotating them to maintain similarity with an upright face (“inverted positions”).

In order to consider the image-based differences introduced by these different flanker stimuli, we computed the Fourier energy for each flanker type in a range of spatial frequency and orientation bands. Analyses were conducted on four faces: the upright target face, an inverted version of the face (used in our experiments as both target and flankers), and the two Thatcherised flanker faces with either inverted features or inverted positions. Images were padded with grey values to extend image dimensions to 1024×1024 pixels. Each image was then fast Fourier transformed before the amplitude spectrum was multiplied with wrapped Gaussian filters with a standard deviation of 20°, centred on orientations between 0° (horizontal) and ±90° (vertical) in 15° increments. Twenty peak spatial frequencies were tested in linear steps between 1-30 cycles per image and a bandwidth of one octave. After the inverse Fourier transform, the RMS contrast of all resulting images was adjusted to match the average RMS contrast of the original image set. Images were then cropped to retain only the central 512×512 pixels to avoid edge artefacts. Example images computed from this stage are shown in Supplementary Figure 1A.

When the energy within the image is summed across all spatial frequencies, it is apparent that each image contained considerably more energy within horizontal bands than within vertical or oblique bands (Supplementary Figure 1B), consistent with previous reports^{7,8}. The overall image content of these faces is therefore highly consistent (particularly since the upright and inverted faces are simply vertically flipped versions of each other, as are the two Thatcherised faces).

To characterise the differences in the target and flanker faces that could drive the crowding of an upright target face (as observed in Experiment 5), we next subtracted each filtered face from the upright filtered face, separately within each spatial frequency and orientation band (on a pixel-by-pixel basis), and again summed the energy across the image. The resulting image differences were then squared and summed to compute the total energy difference between the faces. The average difference (across spatial frequency) within each orientation band is plotted in Supplementary Figure 1C. Firstly, the subtraction of an upright face from itself necessarily produces a zero difference in energy (data not shown). If we instead subtract an inverted face from the upright target (red line in Figure S1C), there are image differences at all orientations, though this is clearly more so in the horizontal bands than in the vertical or oblique ranges. The difference between an upright face and the Thatcherised face with inverted positions follows a nearly identical pattern, with extremely similar values (yellow line in Figure S1C). In contrast, the differences between an upright face and the Thatcherised face with inverted features (in the same positions as the upright face) are considerably reduced at all orientations (purple line in Figure S1C).

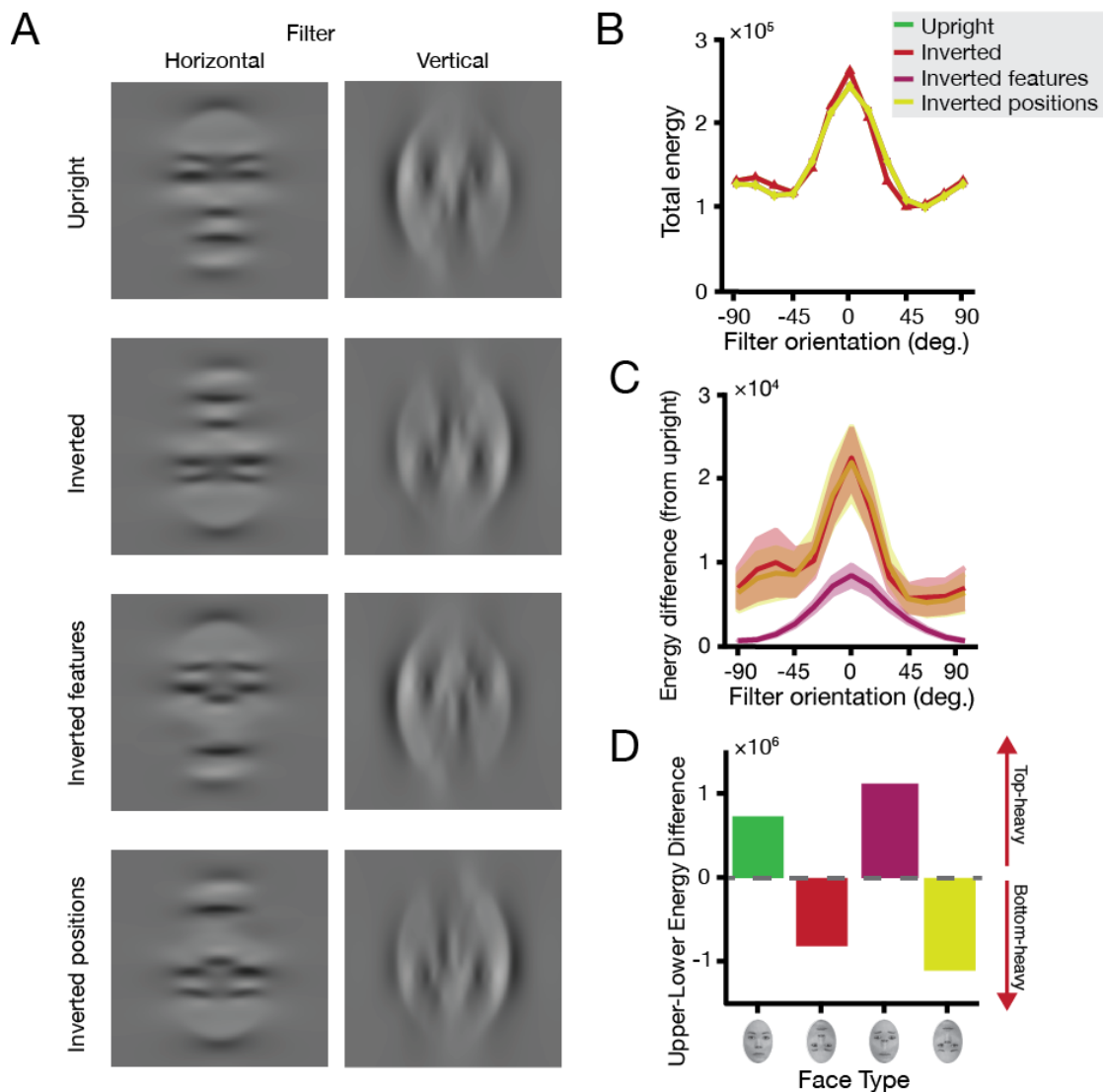


Figure S1: Image analyses for face stimuli

A) Example face stimuli, taken from the Radboud Faces Database⁹ and edited as described in the Methods section of Experiment 5, after convolution with log-Gaussian filters in the Fourier domain. All images are shown after convolution with a peak spatial frequency of 5 cycles per image. The first column shows the following horizontally filtered faces (from top to bottom): an upright face, an inverted face, a Thatcherised face with inverted features, and a Thatcherised face with inverted positions. The second column shows the same faces but vertically filtered.

(B) The total Fourier energy in each image, summed across spatial frequencies within each orientation band (where 0° = horizontal and 90° = vertical). Note that because upright and inverted faces are simply flipped versions of the same stimulus, their values overlap completely. The same is true for the two Thatcherised faces.

(C) Differences in Fourier energy between each face type and an upright face, plotted as a function of the orientation band. Lines show the average and shaded regions show the 95% confidence interval across spatial frequencies. Note that the difference between upright faces lies at zero.

(D) Differences in summed contrast energy between the upper and lower halves of each image used in Experiment 5. Positive values show ‘top-heavy’ and negative values show ‘bottom-heavy’ configurations.

In part, these differences in image structure arise because faces are ‘top-heavy’ stimuli with greater contrast variations in the upper half of the image (primarily due to the eyes and eyebrows) than the lower half^{10,11}. We can quantify this in our stimuli by summing the squared contrast energy at all orientations and spatial frequencies within the lower-half of the image and subtracting this from the sum of squared contrast energy in the upper-half of the image. These values are shown in Figure S1D for the four stimuli used in Experiment 5, where positive values indicate ‘top-heavy’ stimuli and negative values show ‘bottom-heavy’ stimuli. Here it is apparent that an upright face is top-heavy, as is the ‘inverted features’ stimulus with features rotated in place. By contrast, both the inversion of the face image and the inversion of feature positions (with feature orientations kept constant) produces bottom-heavy stimuli.

The image-level differences in Supplementary Figures 1C and 1D show a strong similarity with the results from Experiment 5 (shown in Figure 5). The current analyses show that the inversion of a face produces strong changes in the spatial order of the image, particularly for horizontally oriented structure, with a shift from a ‘top-heavy’ to a ‘bottom-heavy’ configuration. In Experiment 5, this difference was associated with a reduction in crowding with inverted-face flankers, relative to the crowding produced by upright flankers. Similar changes in image content are produced by re-arranging the positions of the upright features to match an inverted face – accordingly, in Experiment 5 we observe a clear reduction in crowding with these flankers. There was far less change in image content when the features were rotated in place, which indeed produced no change in crowding (relative to that produced by upright flankers).

The differences between upright and inverted faces in the distribution of image content have previously been used to examine holistic processes in face recognition, showing that horizontal content is a more effective driver of the behavioural signatures of configural processing^{7,12,13}, and that top-heavy stimulus configurations drive infants’ looking preferences^{10,11}. However, with crowding it is important to note the relations between target and flanker stimuli along these dimensions. Here we show that this image content differs markedly between upright

and inverted faces, as well as with Thatcherised faces with large changes in the spatial ordering of features. These differences in image content could determine the strength of crowding in face stimuli in a similar fashion to the changes produced by other differences in dimensions such as colour¹⁴⁻¹⁷, contrast polarity¹⁸, and orientation¹⁹⁻²². For instance, differences in feature positions of this kind have previously been shown to modulate crowding within letter-like stimuli – an upright T will be crowded less by inverted T flankers than by other configurations²³. We consider these mechanisms further in the General Discussion.

Quantifying target-flanker similarity with Mooney faces

The results of Experiment 5, combined with the above analyses of the orientation energy within face stimuli, suggest that the crowding of face stimuli is driven by target-flanker similarity in the top-heavy vertical configuration of horizontally oriented image structure, similar to the 'bar codes' argued to influence holistic processing in general^{24,25}. However, evidence for an holistic stage of crowding derives not only from the recognition of photographic face stimuli²⁶, but also from the usage of Mooney faces²⁷ where local image content is much more difficult to discern²⁸. As we argue in the General Discussion, although it is true that these stimuli are degraded in terms of the visibility of their local image features (e.g. it is often difficult to make out a nose from an isolated patch of the image), it is not the case that Mooney faces contain no features at all. For instance, Experiment 3 of the study by Farzin et al.²⁷ demonstrates that these stimuli are susceptible to self-crowding between local features in a similar way to photographic images of faces²⁹. Mooney faces also contain a spatial configuration of oriented image structure that is, by and large, similar to that of regular faces²⁴. Here we consider the effect of inversion on this image structure and how this may interact with crowding.

To examine the image content of Mooney faces, an identical set of analyses was performed for a set of Mooney faces as performed on our photographic face stimuli above. A set of 24 Mooney faces was obtained from the Mooney-MF database within the Psychological Image Collection at the University of Stirling*. 12 were male and 12 female, with several faces that match those used previously by Farzin et al.²⁷. To match stimuli to those analysed above, images were reduced to the same dimensions (395×292 pixels), with an oval-shaped aperture with the same dimensions then placed around stimuli to match the image shape required for presentation as crowding stimuli (as in Farzin et al.²⁷). Two example images are shown in Supplementary Figures 2A and 2B.

What is immediately apparent upon examining these Mooney images (and example images in the study of Farzin et al.²⁷) is their increased variability relative to standard face stimuli. In addition to standard frontal views of the face,

* Available online at <http://pics.stir.ac.uk/>

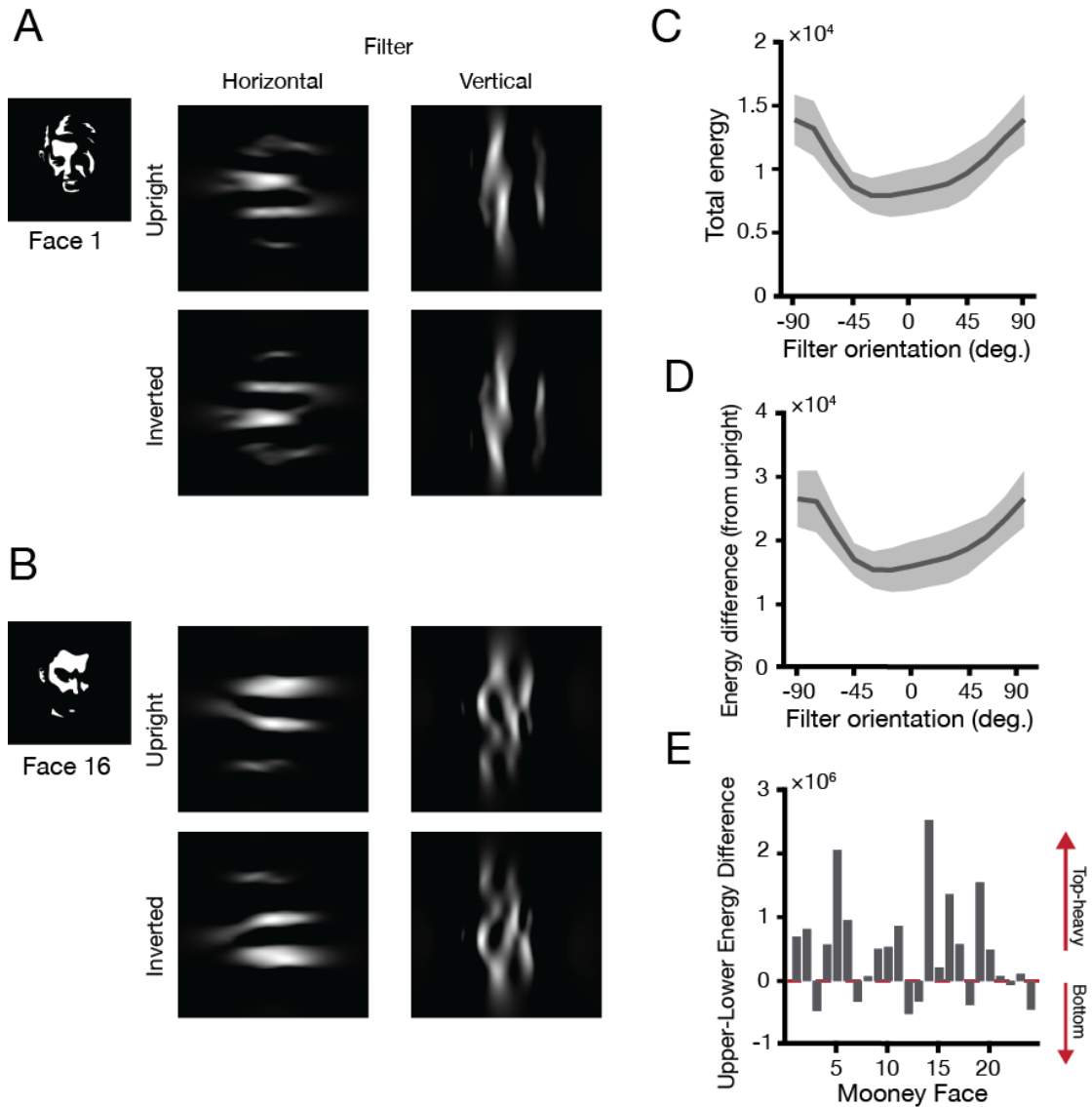


Figure S2. Image analyses for Mooney faces

(A&B). Face 1 (A) and Face 16 (B) from the Stirling Mooney database*. The first column shows the faces after horizontal filtering with a peak spatial frequency of 8 cycles per image. The second column shows the same frequency band after vertical filtering. The first row shows the results for upright Mooney faces and the second for inverted ones.

(C) The mean Fourier energy in each Mooney image, summed across spatial frequencies within each orientation band (where 0° = horizontal and 90° = vertical). Lines show the average and shaded regions show the 95% confidence interval across spatial frequencies.

(D) The mean difference in Fourier energy between upright and inverted versions of each Mooney face, plotted as a function of the orientation band (with conventions as in panel C).

(E) Differences in summed contrast energy between the upper and lower halves of each of the 24 upright Mooney faces. Positive values show ‘top-heavy’ and negative values show ‘bottom-heavy’ configurations.

many Mooney stimuli are in a rotated side profile (as in both examples in

Supplementary Figure 2). The vertical content in these images is often further

*Available online at <http://pics.stir.ac.uk/>

emphasised by their being lit from the side. We therefore expected to see greater variability in the image content.

To examine the orientation energy within these images, the same analyses as above were performed – each image was padded with either black or white brightness values (matching the dominant image background) to extend image dimensions to 1024×1024 pixels. Each image was then fast Fourier transformed before the amplitude spectrum was multiplied with wrapped Gaussian filters with a standard deviation of 20°, centred on orientations between 0° (horizontal) and ±90° (vertical) in 15° increments. Twenty peak spatial frequencies were tested in linear steps between 1-30 cycles per image and a bandwidth of one octave. After the inverse Fourier transform, the RMS contrast of all resulting images was adjusted to match the average RMS contrast of the original image set. Images were then cropped to retain only the central 512×512 pixels to avoid edge artefacts. Inverted images were produced by mirror-reversing the image on the vertical plane. Examples are shown after horizontal and vertical filtering at 8 cycles per image for the two example faces. Notice (especially for the face in Supplementary Figure 2A) that there is a strong vertical component to the image and that the position of this energy changes after image inversion. The same is true for horizontal content, particularly for the face in Supplementary Figure 2B.

The total energy within each face was calculated by summing across all spatial frequencies. The mean value across orientation for the 24 Mooney faces examined herein is shown in Supplementary Figure 2C. Here it can be seen that the average across faces shows a bias towards the vertical orientations, unlike the dominance in horizontal bands with photographic face images. As above, this is due to the many Mooney images that have side-profile views with lighting from the side. We next characterised the differences in potential target and flanker faces that could drive the crowding of an upright target face (as observed in the study by Farzin et al.²⁷). Inverted versions of each Mooney face were filtered and subtracted from upright versions, separately within each spatial frequency and orientation band (on a pixel-by-pixel basis), with the energy across the image then summed. The resulting image differences were squared and summed to compute the total energy difference between the faces. The average difference (across spatial frequency)

within each orientation band is plotted in Supplementary Figure 2D. This analysis reveals image differences between upright and inverted stimuli at all orientations, though unlike the pattern for photographic face images, this is greater in the vertical bands than in other ranges.

Finally, as with the photographic images used in Experiment 5, we also computed the extent to which each Mooney face was top- or bottom-heavy in its distribution of contrast energy. This was computed by summing the squared contrast energy in the bottom half of the image and subtracting this value from the same sum in the upper half. Results are shown in Supplementary Figure 2E for upright versions of each of the 24 faces in this set, where positive values again indicate top-heavy facial configurations. A majority of the upright Mooney faces in this set were indeed top-heavy (17/24), though some showed small biases towards being bottom-heavy. The mean value is nonetheless positive (4.82×10^5), and a one-sided t-test shows this value to be significantly different from zero ($t_{23} = 2.91$, $p = .008$). In other words, when they are upright, Mooney faces are on average top-heavy stimulus configurations just like photographic faces. The inversion of these stimuli would then predominantly produce bottom-heavy stimuli.

The outcome of these analyses suggests that similar processes can account for the effects of crowding on Mooney faces²⁷ as those that we report to account for the effects on photographic face images²⁶. Although Mooney faces are certainly more difficult to recognise than photographic face images, perhaps due to the attenuation of horizontal content shown in our analyses above, it is not the case that they contain no oriented content at all. Our analyses here demonstrate that a bank of oriented filters produces outputs with a bias towards the vertical, and that the structure of these variations changes from a predominantly top-heavy configuration to one that is predominantly bottom-heavy when inverted. As in the results of our Experiment 5, this shift from top- to bottom-heavy configuration in flanker elements (relative to an upright target face) would be expected to reduce crowding, similar to the effects seen with letter-like stimuli²³. As we argue in the General Discussion, these stimuli can certainly produce inter-feature crowding (Farzin et al.²⁷, Experiment 3), suggesting that these contours are themselves susceptible to crowding, just as occurs within photographic face stimuli²⁹. The effects of task

difficulty are also apparent with these stimuli (Farzin et al.²⁷, Experiment 4), just as seen in previous studies with photographic images²⁶ and in Experiment 4 of our study. We therefore argue that Mooney faces do not require an additional stage of holistic crowding to account for the observed effects of flanker orientation – simpler operations based on target-flanker similarity of the kind invoked more generally to account for crowding suffice to account for the entirety of these effects.

Simulations of the effects of task difficulty on face crowding

In Experiment 2 we demonstrate that the selectivity of face crowding is identical for upright and inverted target faces when the task involves judgements of horizontal eye position. This differs from the patterns observed in Experiment 1 for judgements of identity, and in Experiment 3 with judgements of vertical eye position (Figure 3). We attribute this to the susceptibility of these latter tasks to inversion³⁰⁻³², not because this eliminates holistic processing *per se*, but because the resulting increase in task difficulty obscures the selectivity of crowding for target-flanker similarity. Accordingly, the results of Experiment 4a demonstrate that it is possible to obscure the selectivity of crowding with upright faces by increasing task difficulty. We achieved this by reducing the interocular separation in our face stimuli. Conveniently, this also presents the opportunity to conduct model simulations of this process, and to consider how this finding generalises to other tasks like identity judgements.

We thus performed a set of simulations on the effect of crowding for faces, in the context of horizontal eye judgements. This allowed us to consider both the potential mechanisms underlying these crowding effects and the effect of task difficulty. We first assume that there exists a population of detectors that is sensitive to dimensions such as eye position. This is consistent with both theoretical proposals regarding “face space”³³, adaptation effects that shift the perceived eye position within faces^{34,35}, and physiological measurements in the Inferior Temporal (IT) lobe of macaques^{36,37}. Here we simulate a population of detectors selective for interocular eye separation in particular. We do so for the ease of modelling, rather than as a proposal that a population of this nature would be specifically utilised for this purpose – of course, interocular eye separation could be encoded either wholly by or in conjunction with cells selective for other facial properties.

Within this population, we assume that each detector is sensitive to a range of eye separations with a Gaussian tuning function that has a peak sensitivity centred on a particular eye separation, and some sensitivity to nearby values of eye separation, similar to the selectivity of V1 neurons for orientation³⁸, MT/V5 neurons for direction³⁹, and so on. Following the principles of population coding⁴⁰ the

distribution of the resulting population activity would be a Gaussian function centred on the eye separation of the stimulus, with a bandwidth of activity equivalent to the sensitivity bandwidth of the underlying detectors. The perceived value of eye separation could then be read out from this distribution (e.g. as the peak response, or via maximum likelihood estimation).

The relationship between detector sensitivity and the population response means that we can simulate the population response directly as a Gaussian function. We do so here by generating a Gaussian function with a base value of 0.1 and a peak of 1.0, using one free parameter for the standard deviation (to mimic the sensitivity bandwidth of the underlying detectors) and another for the magnitude of Gaussian noise that was added to this response distribution. If we encode the “normal” reference face with an eye-separation value of zero, then the population response to this face will be a Gaussian distribution centred on zero, as shown in Figure S3A. These responses are shown as the average of 720 trials (as in our experiments), generated with an SD of 8 pixels for illustration, and with a comparatively large range in the x-axis of ± 60 pixels to make the population response clear. When the crowded target face is the same as the reference (on “target same” trials) then a similar distribution would arise for this interval. A target face with a large inwards shift would produce a similar Gaussian with a mean located at -20 pixels for the easily detected larger inwards shifts (as in Experiment 2). With this model, we can depict the task as involving a judgement regarding whether the peak of the population response lies on either side of a criterion value – depicted here as a dashed line at -10 pixels (Figure S3A). This is an ideal criterion for the 20 pixel eye shifts, sitting midway between the peak response to either face type. Peaks to the right of this criterion would be classed as the “same” as the reference; those to the left would be classed as “different”. We can therefore simulate the task performed by our observers in this way.

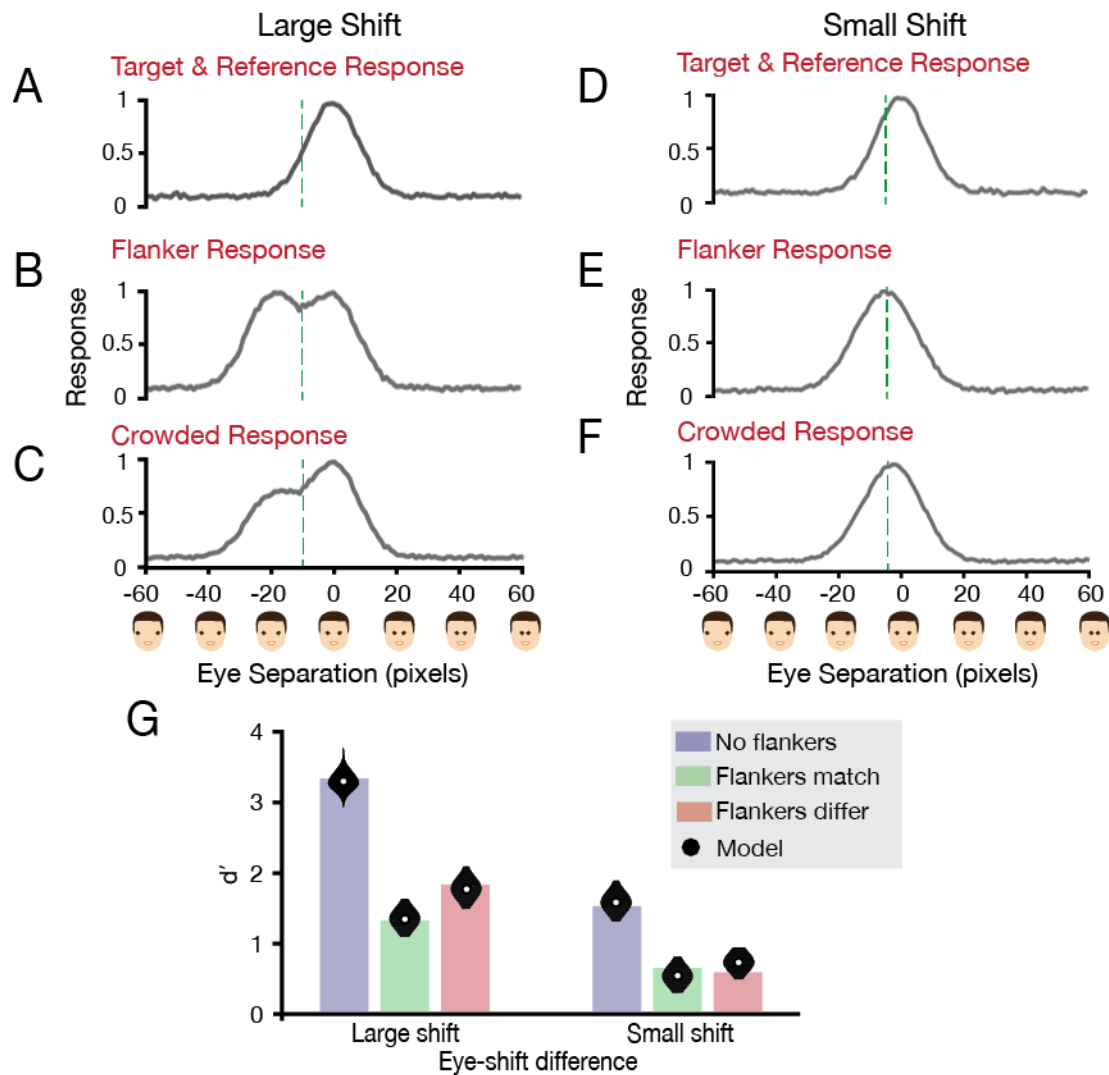


Figure S3: Model simulations for the crowding of eye positions

(A) *Reference response distribution (large shift)*: Population response distributions to the reference face, with interocular eye separation on the x-axis (in pixels) and the population response on the y-axis. The green dashed line indicates the decision criterion at -10 pixels. The solid line plots the mean response across 720 trials (as in our experiments), with the shaded region showing the standard error of the mean. Responses to an unaltered target face (on “target same” trials) would be identical.

(B) *Flanker response distribution (large shift)*: Simulated responses to the combination of all six flanker elements. Three flankers had an eye separation centred on zero (unaltered) and three had the eyes shifted inwards by 20 pixels (as in Experiment 2)

(C) *Crowded response distribution (large shift)*: The combined population response after crowding, modelled as the weighted average of target and flanker responses shown in A and B. A target weight of 0.66 has been applied.

(D) *Reference response distribution (small shift)*: Population response distributions to the reference face with more difficult judgements. Plotted as in panel A, with the decision criterion now at -5 pixels.

(E) *Flanker response distribution (small shift)*: Simulated responses to the six flanker elements. Three flankers had zero eye separation; three had the eyes shifted inwards by 10 pixels (as in Experiment 4a and 4b).

(F) *Crowded response distribution (small shift)*: The combined population response after the weighted average of target and flanker responses in D and B with a target weight of 0.66.

(G) *A comparison of observed and simulated d' values*. Mean d' values from Experiments 2 and 3 are shown here as bars, with simulated values from the best-fitting version of our model shown via violin plots. The simulated mean d' (of 1000 simulations) for each condition is shown as a black circle.

On crowded trials, targets were surrounded by six flankers: three reference faces and three faces with eyes shifted inwards (by 20 pixels in Experiment 2). The combined population response to these flankers would be a bimodal profile with peaks at each of the two eye-separation values, as shown in Figure S3B (again the average of 720 trials). In order to simulate crowding in these instances, we follow recent models of crowding that depict the process as a pooling of target and flanker elements⁴¹⁻⁴⁵, and in particular to models that attribute this pooling to a combination of population responses to the target and flanker elements^{42,43}. Rather than directly averaging these population profiles, we take a weighted average that allows a modulation of the precise combination of target and flanker responses, similar to previous models^{41,46}. When the weighting of the target is high in this combination (relative to that of the flankers) there will be less crowding than when the weighting of the target is low. In this model, the precise weighting of the target was set as a third free parameter ranging from 0-1, with the flanker weighting equal to one minus this value. If we multiply the population response of the target by the target-weighting value, and the population response to the flankers by the flanker-weighting value, then the crowded combination is produced by adding these values.

An example crowded population profile for the large eye-shifts is plotted in Figure S3C, produced with a target weight of 0.66. This gives a bimodal response, albeit with a higher peak (on average) for the target eye separation, given the higher weighting of the target response in this “target same” trial. For this population and this value of the target weight, we observe the reverse pattern of bimodality on target-different trials when the response to the target would be centered on -20 pixels. Nonetheless, in both cases, the secondary peak in this response distribution (caused by the flankers) increases the potential noise in the population response to lead to errors on individual trials. The potential for errors is increased as the flanker weighting increases.

To obtain a release from crowding when the flankers are inverted (as in the “flankers differ” condition of our results), we can therefore simply reduce the weight of the flankers in the weighted average. This is similar to the way that prior models have simulated the effects of target-flanker similarity⁴⁶ and the reduction in

crowding with an increase in target spacing⁴². In other words, with an upright face amongst upright flankers there is a high flanker weight in the average, which is then reduced for an upright target amongst inverted flankers. The precise degree of the release from crowding is the fourth free parameter in our model, implemented as a value between 0-1 that is subtracted from the target weight in the “flankers match” condition.

Task difficulty is introduced here simply by decreasing the eye-separation to 10 pixels, as in Experiment 4a. As with the larger eye shift, the population response distribution to a “normal” reference face would be centered on zero (Figure S3D). When the target matches the reference face, the response would be identical. However, when the target face has a small inward eye shift (“target different” trials), the response would be represented by a Gaussian distribution centered at -10 pixels. The ideal criterion value would thus lie at -5 pixels, sitting midway between the peak response value to a normal face and the peak to a face with a 10 pixel inwards eye shift (dashed line, Figure S3D). The target distribution is then combined with the population response to the flankers. As in Experiment 2, three flankers had “normal” eyes and three had eyes shifted inwards, in this case by 10 pixels. Given the broad response profiles to these values, the reduction in eye shift for this experiment means that the combined response to the six flankers in the “flankers match” condition would have a unimodal profile centred on the criterion value (Figure S3E).

As seen in Figure S3F, the response profile for the combined target and flanker responses (produced with a weight of 0.66) is also unimodal, with a peak between the target value and the decision criterion. As such, there is a higher rate of responses on the side of the flankers compared to the 20 pixel eye shift, which would increase the number of errors. A similar increase in errors would be observed in “target different” trials. This illustrates that when difficulty is increased by reducing the eye shift (as in Experiment 4a), there is greater overlap between target and flanker distributions, and thus a greater propensity for errors to arise due to noise.

From the above distributions we can obtain d' values by simulating both target-same and target-different trials and extracting the peak population response on each trial. Using the location of these peaks on each trial in relation to the

decision criterion (as above), we can score each trial as producing a correct or incorrect response, and then compute a percent correct score in each condition, as in our experiments. We performed these simulations with the same number of trials as the real experiment – 720 trials per observer with 5 observers – repeated 1000 times. The best-fitting parameters for the model were an SD of 13.47 pixels, a noise magnitude of 0.32, target weighting of 0.66 (out of 1), and a crowding release of 0.09 (the difference in target weight in the “flankers match” vs. “flankers differ” conditions).

The results of these simulations are shown in Figure S3G. Mean d' values from our experiment are shown as the bar plots, with simulated distributions of d' values shown in each case as a violin plot for the distribution of all simulated values, where the mean of the 1000 simulations is a white circle. For the “easy” condition with large eye shifts, the model clearly follows the data – d' values are high when the target is uncrowded, decreased with flankers that match the upright orientation of the target, and less impaired when the flankers differ by being inverted. For the difficult condition (“small shift”), uncrowded performance drops significantly, with a further decrease for the crowded conditions. The release from crowding is then considerably muted in these simulations – because performance is reduced overall, the effect of noise is greater and the release from crowding has far less effect.

The mean difference between crowded d' values for our model is 0.42 with large eye shifts and 0.20 for the more difficult small eye-shift condition. Note that the effect of task difficulty in this sense has nothing to do with our free parameters – this is simply introduced by altering the input values for eye separation from -20 to -10 in the target-different conditions. It is the combination of lowered performance and noise that flattens the selectivity of crowding for target-flanker similarity.

How might we then implement the effect of inversion within this framework? For example, for the effects we observe in Experiment 3 with vertical eye-judgements, inversion could be implemented in several ways. Inversion is generally thought to produce a shift from configural processing to more local processing¹. This may result in either a specific impairment in configural dimensions (that are not coded in a local/featural manner) or the use of inappropriate facial landmarks. To model the effects of vertical eye-position in this sense, inversion could be

implemented as an increase in the noise associated with these eye-judgements, thereby increasing the propensity for errors, or it may broaden either the spatial or featural selectivity of the detectors sensitive to eye position (as suggested for the orientation selectivity of face identification⁸, for instance). In these cases, the effects of inversion would be similar to the effect of reduced eye displacements modeled above.

Of course, inversion also disrupts the identification of faces^{31,32}. These effects could similarly be modelled in a crowding paradigm via an increase in noise, an increase in the spatial or featural bandwidth of the underlying detectors, or even as a shift in the population responses towards the decision boundary. In this case, the population response would be distributed across dimensions more relevant to identity (e.g. within a “face space”³³). Nonetheless, by expanding out from our simple model of interocular separation, we can consider that the effect of crowding on judgements of identity may arise in a similar fashion to that observed herein with judgements of eye position. Importantly however, it is not the mechanisms of crowding that would change with inversion in these cases, but rather the difficulty of the task, which in turn determines whether the selectivity of crowding for target-flanker similarity is evident or not. In this sense, we argue that crowding shares a common mechanism in all cases, rather than requiring processes specific to the holistic encoding of faces.

Supplemental References

- 1 Rossion, B. Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica* **128**, 274-289 (2008).
- 2 Tanaka, J. W. & Farah, M. J. Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: A Human Experimental Psychology* **46**, 225-245 (1993).
- 3 Young, A. W., Hellawell, D. & Hay, D. C. Configural information in face perception. *Acta Psychologica* **128** (1987).
- 4 Le Grand, R., Mondloch, C. J., Mauer, D. & Brent, H. P. Neuroperception: Early visual experience and face processing. *Nature* **410**, 890 (2001).
- 5 Diamond, R. & Carey, S. Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General* **115**, 107-117 (1986).
- 6 Thompson, P. Margaret Thatcher: A new illusion. *Perception* **9**, 483-484 (1980).
- 7 Dakin, S. C. & Watt, R. J. Biological "bar codes" in human faces. *Journal of Vision* **9**, 2 (2009).
- 8 Goffaux, V. & Greenwood, J. A. The orientation selectivity of face identification. *Scientific Reports* **6**, 34204 (2016).
- 9 Langner, O. *et al.* Presentation and validation of the Radboud Faces. *Cognition & Emotion* **24**, 1377-1388 (2010).
- 10 Viola Macchi, C., Turati, C. & Simion, F. Can a nonspecific bias towards top-heavy patterns explain newborns' face preference? *Psychological Science* **15**, 379-383 (2004).
- 11 Simion, F., Valenza, E., Macchi Cassia, V., Turati, C. & Umiltà, C. A. Newborn's preference for up-down asymmetrical configurations. *Developmental Science* **5**, 427-434 (2002).
- 12 de Heering, A. *et al.* Three-month-old infants' sensitivity to horizontal information within faces. *Developmental Psychobiology* **58**, 536-542 (2016).
- 13 Glen, J. C. & Dakin, S. C. Orientation-crowding within contours. *Journal of Vision* **13**, 1-11 (2013).
- 14 Pöder, E. Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research* **71**, 641-645 (2007).
- 15 Kennedy, G. J. & Whitaker, D. The chromatic selectivity of visual crowding. *Journal of Vision* **10**, 15 (2010).
- 16 Kooi, F. L., Toet, A., Tripathy, S. P. & Levi, D. M. The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision* **8**, 255-279 (1994).
- 17 Gheri, C., Morgan, M. J. & Solomon, J. A. The relationship between search efficiency and crowding. *Perception* **36**, 1779-1787 (2007).
- 18 Chung, S. T., Levi, D. M. & Legge, G. E. Spatial frequency and contrast properties of crowding. *Vision Research* **41**, 1833 - 1850 (2001).
- 19 Andriessen, J. J. & Bouma, H. Eccentric vision: Adverse interactions between line segments. *Vision Research* **16**, 71-78 (1976).
- 20 Wilkinson, F., Wilson, H. R. & Ellemberg, D. Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A* **14**, 2057-2068 (1997).

- 21 Hariharan, S., Levi, D. M. & Klein, S. A. "Crowding" in normal and amblyopic vision assessed with Gaussian and Gabor C's. *Vision Research* **45**, 617-633 (2005).
- 22 Levi, D. M., Hariharan, S. & Klein, S. A. Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant or simple contrast masking. *Journal of Vision* **2**, 3 (2002).
- 23 Dakin, S. C., Cass, J., Greenwood, J. A. & Bex, P. J. Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *J Vis* **10**, 14, doi:10.1167/10.10.14 (2010).
- 24 Dakin, S. C. & Watt, R. J. Biological "bar codes" in human faces. *Journal of Vision* **9**, 1-10 (2009).
- 25 Goffaux, V. & Dakin, S. C. Horizontal information drives the behavioural signatures of face processing. *Frontiers in Psychology* **1**, 143 (2010).
- 26 Louie, E. G., Bressler, D. W. & Whitney, D. Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision* **7**, 24-24 (2007).
- 27 Farzin, F., Rivera, S. M. & Whitney, D. Holistic crowding of Mooney faces. *Journal of Vision* **9**, 1-15 (2009).
- 28 Mooney, C. Age in the development of closure ability in children. *Canadian Journal of Psychology* **11**, 219-226 (1957).
- 29 Martelli, M., Majaj, N. J. & Pelli, D. G. Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision* **5**, 58-70 (2005).
- 30 Goffaux, V. & Rossion, B. Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception & Performance* **33**, 995-1001 (2007).
- 31 Yin, R. K. Looking at upside-down faces. *Journal of Experimental Psychology* **81**, 141-145 (1969).
- 32 McKone, E. Isolating the special component of face recognition: Peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 181-197 (2004).
- 33 Valentine, T. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology* **43**, 161-204 (1991).
- 34 Robbins, R., McKone, E. & Edwards, M. Aftereffects for face attributes with different natural variability: Adapter position effects and neural models. *Journal of Experimental Psychology: Human Perception & Performance* **33**, 570-592 (2007).
- 35 Susilo, T., McKone, E. & Edwards, M. What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research* **50**, 300-314 (2010).
- 36 Freiwald, W. A., Tsao, D. Y. & Livingstone, M. S. A face feature space in the macaque temporal lobe. *Nature Neuroscience* **12**, 1187-1196 (2009).
- 37 Chang, L. & Tsao, D. Y. The code for facial identity in the primate brain. *Cell* **169**, 1013-1028 (2017).

- 38 Schiller, P. H., Finlay, B. L. & Volman, S. F. Quantitative studies of single-cell properties in monkey striate cortex | |. Orientation specificity and ocular dominance. *Journal of Neurophysiology* **38**, 1320-1333 (1976).
- 39 Felleman, D. J. & Kaas, J. H. Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology* **52**, 488-513 (1984).
- 40 Pouget, A., Dayan, P. & Zemel, R. Information processing with population codes. *Nature Reviews Neuroscience* **1** (2000).
- 41 Greenwood, J. A., Bex, P. J. & Dakin, S. C. Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12130-13135 (2009).
- 42 Harrison, W. J. & Bex, P. J. A unifying model of orientation crowding in peripheral vision. *Current Biology* **25**, 3213-3219 (2015).
- 43 van den Berg, R., Roerdink, J. B. T. M. & Cornelissen, F. W. A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology* **6**, e1000646 (2010).
- 44 Parkes, L., Lund, J., Angellucci, A., Solomon, J. A. & Morgan, M. Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience* **4**, 6 (2001).
- 45 Freeman, J., Chakravarthi, R. & Pelli, G. D. Substitution and pooling in crowding. *Attention, Perception, & Psychophysics* **74**, 379-396 (2012).
- 46 Greenwood, J. A., Bex, P. J. & Dakin, S. C. Crowding follows the binding of relative position and orientation. *Journal of Vision* **12**, 1-20 (2012).