OPEN ACCESS

University of BRISTOL

Rhodes, G., Pond, S., Jeffery, L., Benton, C., Skinner, A., & Burton, N. (2016). Aftereffects Support Opponent Coding of Expression. Journal of Experimental Psychology: Human Perception and Performance.

Peer reviewed version

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) will be available online via APA. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms.html

1

Aftereffects Support Opponent Coding of Expression

Gillian Rhodes[1], Stephen Pond[1], Linda Jeffery[1], Christopher P. Benton[2], Andrew L.

Skinner[2,3] & Nichola Burton[1]

[1]ARC Centre of Excellence in Cognition and its Disorders, School of Psychology,

University of Western Australia, Australia

[2] School of Experimental Psychology, University of Bristol, United Kingdom & ARC

Centre of Excellence in Cognition and its Disorders

[3] Medical Research Council Integrative Epidemiology Unit at the University of

Bristol, United Kingdom

WORD COUNT: 5234 excluding references and tables

Please address correspondence to:
Professor Gillian Rhodes
School of Psychology, University of Western Australia
35 Stirling Highway, Crawley, WA 6009
AUSTRALIA
phone: +61-8-6488-3251
fax: +61-8-6488-1006
email: gillian.rhodes@uwa.edu.au

Running head: Expression coding

Keywords: Face perception, expression perception, expression aftereffects, opponent

coding

Acknowledgements

Abstract

We used aftereffects to investigate the coding mechanisms underlying our perception of facial expression. Recent evidence for dimensions that are common to the coding of both expression and identity suggest that the same coding system could be used for both attributes. Identity is adaptively opponent coded by pairs of neural populations tuned to opposite extremes of relevant dimensions. Therefore, we hypothesized that expression would also be opponent coded. An important line of support for opponent coding is that aftereffects increase with adaptor extremity (distance from an average test face) over the full natural range of possible faces. Previous studies have reported that expression aftereffects increase with adaptor extremity. Critically, however, they did not establish the extent of the natural range and so have not ruled out a decrease within that range that could indicate narrowband, multichannel coding. Here we show that expression aftereffects, like identity aftereffects, increase linearly over the full natural range of possible faces and remain high even for impossibly distorted adaptors. These results suggest that facial expression, like face identity, is opponent coded.

Aftereffects Support Opponent Coding of Expression

There has been considerable interest in the coding systems underlying our expertise in face perception. The concept of a multidimensional face-space, in which faces are mentally represented on some set of perceptual dimensions, has provided an influential framework for understanding many aspects of face processing (Rhodes, 1988; Rhodes, Brennan, & Carey, 1987; Valentine, 1986, 1991, 1995). However, distinct computational coding systems are possible within this framework, and different systems are possible for different face attributes. Here we focus on the coding of facial expression. Our ability to read facial expressions is critical for social interaction, and requires the discrimination of subtle variations between facial configurations. How might this be done?

Despite the traditional focus on distinct pathways for coding of expression, a changeable attribute, and identity, a more stable attribute (Bruce & Young, 1986; Haxby & Gobbini, 2011; Haxby, Hoffman, & Gobbini, 2000, 2002; Hoffman & Haxby, 2000), there is now evidence for common coding at the level of perceptual representations. The selectivity of visual neural processing for these attributes is far from complete and the classic neuropsychological dissociation between deficits in identity and expression recognition may arise post-perceptually (for a review see Calder, 2011). In addition, Principal Components Analysis (PCA) shows that common image components (cf. dimensions) can support the discrimination of both identity and expression (Calder, Burton, Miller, Young, & Akamatsu, 2001) and in humans there are common dimensions that contribute to the recognition of both attributes. Specifically, the adaptive coding of expression and identity (measured by aftereffects) shares common variance, which significantly predicts recognition of both

attributes (Rhodes et al., 2015). The common dimensions could include spatial relations that vary with both identity and expression (e.g., eyebrow height - low for anger, high for surprise), feature attributes that vary with identity and expression (e.g., lip thickness - decreases for anger), and/or holistic dimensions, like PCA eigenfaces that can represent both attributes (Calder et al., 2001).

Given that common dimensions contribute to the coding of both expression and identity, we propose that expression is coded using the same type of computational mechanisms as identity. Identity is adaptively coded relative to norms that are updated by experience (for reviews see Rhodes & Leopold, 2011; Webster & MacLeod, 2011). This norm-based coding of identity-related dimensions appears to be implemented by opponent coding, with pairs of neural populations tuned to opposite extremes of each dimension and equal activation in the two populations implicitly signalling the norm (Fiorentini, Gray, Rhodes, Jeffery, & Pellicano, 2012; Jeffery, Read, & Rhodes, 2013; Jeffery et al., 2011; McKone, Jeffery, Boeing, Clifford, & Rhodes, 2014, 2015; Robbins, McKone, & Edwards, 2007; Susilo, McKone, & Edwards, 2010). These studies show that identity-related aftereffects increase with adaptor extremity, as predicted by opponent coding.[1] This monotonic pattern is predicted because more extreme adaptors activate their preferred channel more strongly (and their non-preferred channel more weakly) than less extreme adaptors, producing a stronger reduction in response and therefore a larger aftereffect. This prediction has been confirmed by quantitative modelling (McKone et al., 2014).

These studies provided no support for an alternative non-norm-based coding system, narrowband multichannel coding, which is used for several basic visual attributes, including tilt and spatial frequency (Blakemore & Sutton, 1969; Clifford,

---

[1] McKone et al (2014) initially reported a non-monotonic pattern, but this was due to an error in a single data point, which was corrected in McKone et al (2015).

Wenderoth, & Spehar, 2000). That coding model predicts a non-monotonic pattern of initial increase in aftereffects followed by a decrease as increasingly extreme adaptors have less and less impact on channels that respond to the (average) test face (McKone et al., 2014). Again, this predicted pattern has been confirmed by quantitative modelling (McKone et al., 2014). Moreover, there were strong identity-related aftereffects for impossibly extreme adaptors, which should have little impact on responses of channels tuned (narrowly) to the (average) test faces used (McKone et al., 2014, 2015; Robbins et al., 2007; Susilo et al., 2010). Finally, for identity-related features such as eye height, there was no generalized repulsion away from the adaptor level, as predicted by narrowband multichannel coding, but rather a uniform shift in the whole response curve, consistent with renormalization (Robbins et al., 2007).

A recent paper has argued against opponent coding for face identity (Storrs & Arnold, 2015), demonstrating local repulsion as expected from multichannel coding, rather than a consistent renormalization as expected from opponent coding. They used a spatial comparison task, which can only measure the spatiotopic component of face aftereffects. However, face perception and face aftereffects, as normally measured, clearly have global, non-spatiotopic components (S.-R. Afraz & Cavanagh, 2008; Leopold, O'Toole, Vetter, & Blanz, 2001) and indeed some aftereffects (face gender) have no spatiotopic component at all (A. Afraz & Cavanagh, 2009). Therefore, we suggest that their results are not informative about higher-level identity coding mechanisms.

To demonstrate opponent coding, it is critical that aftereffects increase over the full natural range of possible faces. Identity aftereffects show exactly this pattern, and remain high even for impossibly distorted adaptors (McKone et al., 2014, 2015). A few studies have reported that expression aftereffects also increase with adaptor

extremity, consistent with opponent coding (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Skinner & Benton, 2010, 2012). Critically, however, they have not established the extent of the natural range, and so cannot rule out a decrease within that range that would indicate narrowband multichannel coding. Moreover, these studies sampled adaptation strength rather sparsely and the pattern of increase was not entirely consistent, with aftereffects increasing from 50% to 100% adaptor levels in one study (Skinner & Benton, 2010) and from 25% to 50%, but not from 50% to 100%, adaptor levels, in another (Skinner & Benton, 2012). (These adaptors are anti-expressions created by morphing target expressions towards, and beyond, an average expression. A 100% adaptor is equally distant from the average (in morph steps) as the original expression).

In the present study, we measured expression aftereffects for adaptors that spanned and exceeded the full natural range of possible faces. Following previous studies, we measured aftereffects as the shift in perception of an average-expression test face towards the expression opposite to the adapting expression (e.g., Skinner & Benton, 2010). For example, adapting to an anti-happy expression (made by caricaturing an average expression away from a happy expression) should bias perception towards a happy expression. We also established the boundary between faces that were perceived as physically possible and impossible, and included a similarity task (rating similarity of adaptors to the average test face) to explicitly check that our adaptors increased in perceived extremity. We hypothesized that expression aftereffects would show the same pattern as identity aftereffects, with a monotonic increase over the full natural range and substantial aftereffects for highly distorted adaptors lying outside this range. We minimized the contribution of low-

level retinotopic adaptation by using a size change between adapt and test faces and allowing free eye movements.

Method

Participants

Thirty-five Caucasian adults (11 male) participated for either course credit or $20 (Mean age = 22.1 years, SD = 6.6 years). All were recruited from the University of Western Australia. We chose the sample size to match those used to measure how other face aftereffects change with adaptor extremity (e.g., McKone et al., 2014; Pond et al., 2013). Two participants failed to return for the second session and were excluded, giving a final sample of 33 (11 male, Mean age = 21.9 years, SD = 6.6 years).

General Procedure

Participants completed two 50-minute sessions. Each session contained an expression adaptation task used to measure expression aftereffects, a natural boundary task used to establish the natural range of possible expressions, and a similarity task designed to check whether morphing expressions further away from the average expression increases the perceived extremity of the resulting anti-expressions, as assumed. The natural boundary task was always completed before the similarity task, and these two tasks preceded the adaptation task in session one and followed it in session two.

Expression Adaptation Task

This task measured expression aftereffects for a range of adaptor extremities within and beyond the natural boundary of physically possible faces. For each of four highly discriminable target expressions (100% expression strength), we created adaptor anti-expressions with varying levels of extremity (physical deviation from the

average: 0, 40, 80, 120, 160, 200, 240, 280, 320, 360, 400, 440%) (Figure 1).

Participants were assigned two expressions (either angry and happy or disgusted and sad) and saw only adaptor faces derived from those expressions. Their task was to decide if an average-expression (0%) test face showed an angry or happy expression (or a disgusted or sad expression). We measured expression aftereffects as the bias to perceive the expression opposite the adapting anti-expression.

Stimuli. The stimuli were constructed from four expression prototypes (angry, happy, disgusted, sad) (Figure 1) and an average expression, taken from Skinner and Benton (2010). Each expression prototype was the average of 50 front-view images of young Caucasian adults (25 male, 25 female) displaying that expression. The average expression was the average of seven expression prototypes (angry, happy, disgusted, sad, surprise, fear, neutral). All prototypes were created using standard morphing procedures.
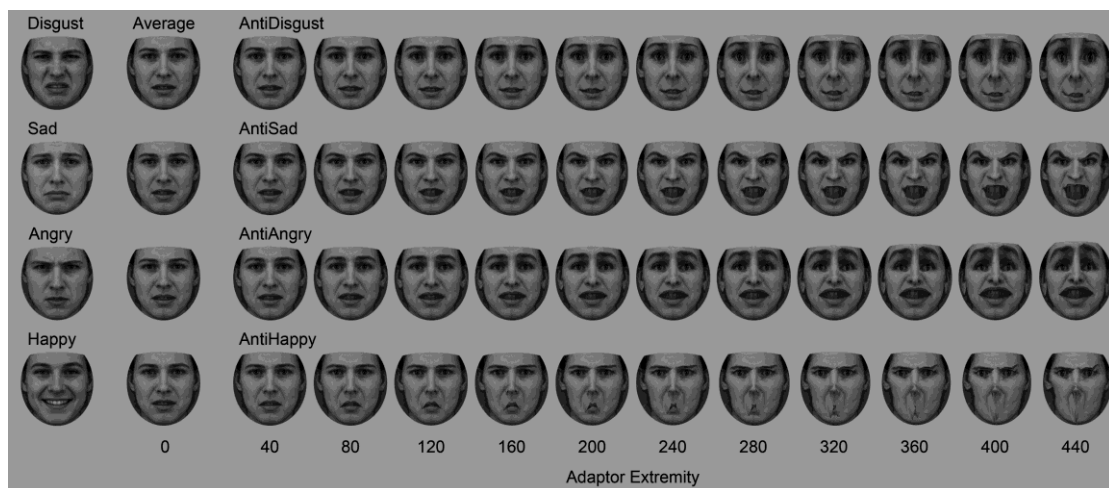

Figure 1. Four anti-expression adaptation continua, each consisting of 12 extremity levels (0, 40, 80, 120, 160, 200, 240, 280, 320, 360, 400, 440%). The 0% adaptor extremity was an average expression prototype and was identical in all continua. The

four target expressions used to make the anti-expression continua are shown on the left.

For each of four target expressions (angry, happy, disgusted, sad), we made anti-expressions at increasing extremity levels (0, 40, 80, 120, 160, 200, 240, 280, 320, 360, 400, 440%) (Figure 1), by caricaturing the average expression (0%) away from the target expression prototype, using standard morphing procedures in Fantamorph 5.3.2. Note that a 100% anti-expression (e.g., 100% anti-happy) lies physically as far from the average (but in the opposite direction) as the corresponding original expression (e.g., 100% happy). The anti-expressions were used as adapting images.

We also made reduced strength versions of the target expressions (40%, 60%, and 80%), by morphing each target expression (angry, happy, disgusted, sad) towards the average expression (0%) using Fantamorph 5.3.2. The primary test image was the average expression (0%), but 80% test images were shown on a minority of trials (see below), to provide some easy trials to help maintain motivation and to confirm that participants could accurately identify "strong" versions of the targets. The 40% and 60% images were used only in training.

Procedure. The adaptation task was identical in each session and data were pooled across the two sessions. It was presented on an iMac with a 20-inch LCD screen, with anti-glare covering, using Cedrus Superlab 4.07 software (Abboud, Schultz, & Zeitlin, 2008). Participants began with training in judging the target expressions, followed by the adaptation trials. Each participant saw stimuli derived from two expressions, either angry and happy or disgusted and sad. For ease of

exposition, we describe training and adaptation procedures for the angry/happy condition.

In training, participants were told to press the "angry" key whenever the test face appeared angry and the "happy" key whenever it appeared happy. Both strong (100%) and weak (40%, 60%) versions of each target were shown, so that participants understood how to respond to weak impressions of expression. Training was split into two stages and took approximately 2 minutes. In the first stage, targets (2 expressions x 3 strengths x 2 repeats, random order) were displayed until a response was made, and in the second stage, targets were shown for 400 ms (as in the adaptation task). Participants received auditory feedback (a bell tone for correct responses and a buzzer noise for incorrect responses) on each trial. All participants scored 10 out of 12 correct or better in both stages.

After training, participants completed the adaptation trials, which took approximately 30 minutes. On each trial participants saw an adapting anti-expression for 5000ms, a 150ms inter-stimulus interval, a test face for 400ms, and a blank gray screen that remained until participants responded. Participants initiated each trial by pressing the spacebar. There were 288 trials in each session, presented in a different random order for each participant. The test face was the average expression (0%) on 240 trials: 10 trials for each adapting expression (anti-angry/anti-happy) at each adaptor extremity (0, 40, 80, 120, 160, 200, 240, 280, 320, 360, 400, 440%). The test face was the 80% target on 48 trials: one trial for each target (angry/happy) shown with each adapting face at each adaptor strength. Trials were split into six blocks and participants were encouraged to take breaks if needed. Participants were shown a cartoon or joke between blocks, which remained visible until they pressed the spacebar to continue, to ensure that a minimal break was taken. The task began with

two practice trials, one for each 80% test face shown after its 160% anti-expression.

We minimized the contribution of low-level (retinotopic) adaptation by using

adapting stimuli (7.6° x 7.6°, viewed from 50 cm) that were larger than the test

stimuli (5.6° x 5.6°) and allowing free eye movements.

Natural Boundary Task

For each anti-expression continuum (2) of 12 images (0, 40, 80, 120, 160, 200,

240, 280, 320, 360, 400, 440%) used in their adaptation task, participants were asked

to indicate the point of switch between being "a normal face (i.e., one that could occur

in the real world), to a distorted face that could not normally occur". Each continuum

was shown twice, once with extremity increasing from left to right and once with

extremity decreasing from left to right (4 trials). Each continuum remained visible

until the participant responded. Responses were made verbally and recorded by the

experimenter. Each face subtended approximately 4.1º x 4.1º viewed from 50 cm,

with approximately 0.4º between adjacent faces. Order of expression continuum

(anti-angry or anti-happy presented first) and extremity direction (increasing or

decreasing strength presented first) were counterbalanced across participants.

Continuum expressions alternated, with the same extremity order in the first two and

second two trials. The task lasted approximately 5 minutes.

Similarity Ratings

Participants rated the similarity of anti-expressions seen in the adaptation task

to the average (0%) expression on a 10-point scale using labelled (1-10) keyboard

keys. They were encouraged to use the full range. On each trial an anti-expression

was presented beside the average (0%) expression, and participants were asked to rate

how similar the two faces looked. Participants rated all eleven anti-expressions (40,

80, 120, 160, 200, 240, 280, 320, 360, 400, 440%) for each of their assigned

expressions (anti-angry and anti-happy or anti-disgusted and anti-sad) twice, once with the anti-expression to the left of the average expression and once to the right, for a total of 44 randomly ordered trials. Each face was shown at the same size as the adapting faces in the adaptation task and subtended a visual angle of 7.6° x 7.6° when viewed from approximately 50cm. Faces were separated by a distance of approximately 7.4°. Participants initiated each trial by pressing the spacebar, and each face pair remained visible until the participant responded. The pairs were presented in random order, and the rating scale remained on screen during all trials. At the beginning of each session, one highly similar pair and one highly different pair were given as examples along with text indicating that the pair should get a high or low rating.

Results

Aftereffects were calculated for each adaptor extremity using trials with the average-expression test face. For happy/angry pairs, the proportion of "angry" responses after anti-happy (mismatch) adaptors was subtracted from the proportion of "angry" responses after anti-angry (match) adaptors (Table 1). For disgust/sad pairs, the proportion of "sad" responses after anti-disgust (mismatch) adaptors was subtracted from the proportion of "sad" responses after anti-sad (match) adaptors (Table 1). This procedure produces positive scores for aftereffects in the predicted direction independent of any bias to choose one response over the other. For 0% adaptors, there was no difference between anti-happy and anti-angry (or anti-disgust and anti-sad) adaptors, and these were arbitrarily dummy-coded either anti-angry or anti-happy (or anti-disgust and anti-sad) in order to calculate an aftereffect (expected to be zero) for these stimuli (as in Leopold et al., 2001).

Disgust-Sad Condition

Adaptor Strength (%)	Disgust Responses: Match (Adapt Anti-disgust)		Disgust Responses: Mismatch (Adapt Anti-sadness)		Aftereffect (Match-Mismatch)	
	M	(SE)	M	(SE)	M	(SE)
0	**0.44**	**(0.07)**	**0.47**	**(0.06)**	-0.03	(0.04)
40	**0.45**	**(0.05)**	**0.41**	**(0.07)**	0.04	(0.04)
80	**0.58**	**(0.05)**	**0.37**	**(0.06)**	0.21	(0.04)
120	**0.64**	**(0.05)**	**0.27**	**(0.05)**	0.37	(0.05)
160	**0.70**	**(0.06)**	0.30	(0.05)	0.40	(0.06)
200	0.60	(0.06)	0.30	(0.06)	0.30	(0.07)
240	0.66	(0.06)	0.28	(0.05)	0.38	(0.07)
280	0.66	(0.06)	0.30	(0.06)	0.36	(0.07)
320	0.63	(0.06)	0.33	(0.06)	0.31	(0.05)
360	0.65	(0.06)	0.32	(0.05)	0.33	(0.06)
400	0.61	(0.06)	0.31	(0.05)	0.30	(0.06)
440	0.59	(0.06)	0.32	(0.06)	0.28	(0.06)

Happy-Angry Condition

Adaptor Strength (%)	Happy Responses: Match (Adapt Anti-happy)		Happy Responses: Mismatch (Adapt Anti-anger)		Aftereffect (Match-Mismatch)	
	M	(SE)	M	(SE)	M	(SE)
0	**0.43**	**(0.07)**	**0.44**	**(0.06)**	-0.01	(0.03)
40	**0.56**	**(0.06)**	**0.40**	**(0.06)**	0.15	(0.05)
80	**0.67**	**(0.06)**	**0.41**	**(0.07)**	0.26	(0.06)
120	0.66	(0.06)	**0.38**	**(0.06)**	0.28	(0.06)
160	0.62	(0.06)	0.39	(0.07)	0.23	(0.07)
200	0.64	(0.06)	0.41	(0.06)	0.22	(0.05)
240	0.59	(0.07)	0.36	(0.06)	0.23	(0.06)
280	0.54	(0.07)	0.38	(0.06)	0.16	(0.04)
320	0.51	(0.07)	0.35	(0.06)	0.16	(0.06)
360	0.51	(0.07)	0.36	(0.07)	0.15	(0.06)
400	0.49	(0.07)	0.38	(0.07)	0.11	(0.05)
440	0.47	(0.08)	0.34	(0.07)	0.13	(0.05)

Table 1. Mean responses used to calculate aftereffects in the Disgust-Sad and Happy-Angry conditions. Bold values are responses following adaptation within the natural boundary for each expression: Adaptors beyond this boundary were judged to appear physically impossible. Aftereffect is calculated by subtracting mismatch responses from match responses, giving an unbiased estimate of the aftereffect.

We also calculated the natural boundary between possible and impossible faces, to establish the extent of the natural range over which aftereffects should increase if expression is opponent coded. Critically, any decrease within this range, would be consistent with narrowband, multichannel coding. The number of participants who rated each morph level as impossible for each expression continuum, (averaged across session and order conditions) is shown in Table 2. The boundary.for each adaptation expression condition was calculated by averaging boundary estimates for the two relevant continua (happy/angry pairs: M = 123, SD = 53, 95%CI = 98, 148; disgust/sad pairs: M = 163, SD = 72, 95%CI = 127, 198) (see Figure 2).

Expression	N	Anti-Expression Morph Level (%)											
		0	40	80	120	160	200	240	280	320	360	400	440
Angry	17	0	0	1	4	9	15	17	17	17	17	17	17
Happy	17	0	0	4	13	17	17	17	17	17	17	17	17
Disgust	16	0	0	0	2	6	11	13	15	16	16	16	16
Sad	16	0	0	2	5	10	12	14	15	16	16	16	16

Table 2. Cumulative numbers of participants who judged each morph level to be "a distorted face that could not normally occur".

Expression Aftereffects

We conducted a two-way repeated measures ANOVA on the aftereffects, with adaptor extremity (0, 40, 80, 120, 160, 200, 240, 280, 320, 360, 400, 440) as a repeated measures factor and expression pair (disgust/sad, happy/angry) as a between-participants factor. Greenhouse-Geisser correction was used where appropriate. There was a significant effect of adaptor extremity, $F(6.40, 198.47) = 13.40$, $p <$ .0001, $\eta_p^2 = .302$, which interacted with expression pair, $F(6.40, 198.47) = 4.45 <$ .0001, $\eta_p^2 = .126$ (Figure 2). There was no main effect of expression pair, $F(1,31) =$

2.80, $p = .104$, $\eta_p^2 = .083$. To explore the interaction, we conducted separate analyses

each expression pair.

Disgust/sad aftereffects. Inspection of Figure 2 shows that these aftereffects

increased numerically up to a peak at 160%, which was close to the mean boundary

for the natural range (M = 163). Over the natural range (0-160), polynomial contrasts

revealed only a significant linear effect, $F(1,15) = 35.03$, $p < .0001$, $\eta_p^2 = .700$.

Planned t-tests showed a marginally significant increase from 0 to 40 ($p = .052$), a

significant increase from 40 to 80 ($p < .007$), and 80 to 120 ($p < .002$), and no

significant increase from 120-160 ($p = .569$). These results clearly support opponent
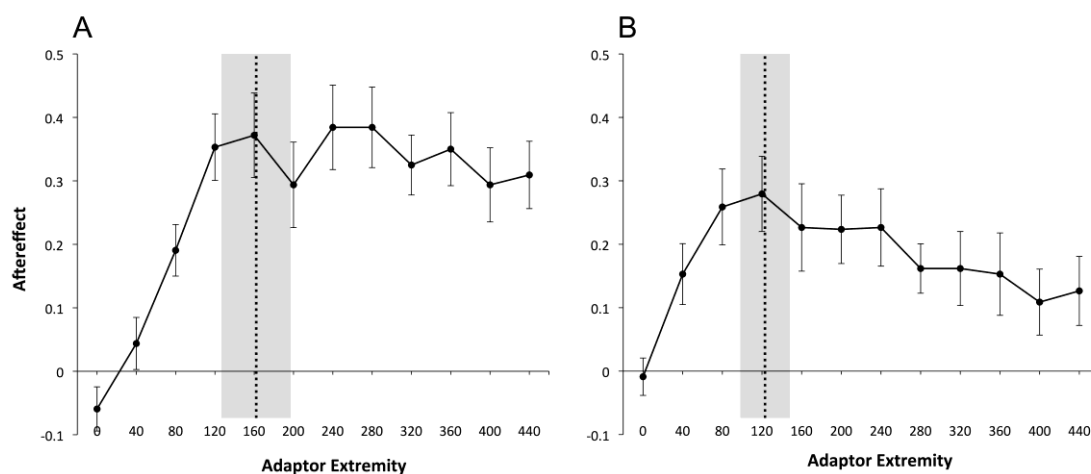
coding of expression.

Figure 2. Expression aftereffects as a function of adaptor extremity for disgust/sad

(A) and angry/happy (B) expression pairs. SE bars are shown. The dotted vertical

lines show the natural boundary between possible and impossible faces for each

expression-pair condition. The shaded grey areas represent 95% confidence intervals

for the natural boundary estimates.

Beyond the natural range (200-440), aftereffects remained high (Figure 2). Indeed, planned comparisons showed that there was no significant decrease, whether or not comparisons were Bonferroni corrected ($ps > .14$, uncorrected). Therefore, impossibly distorted faces can strongly activate the expression coding system.

Over the full range (0-440), polynomial contrasts revealed significant linear, $F(1,15) = 27.14$, $p < .0001$, $\eta_p^2 = .644$, quadratic, $F(1,15) = 34.15$, $p < .0001$, $\eta_p^2 = .695$, and cubic effects, $F(1,15) = 9.02$, $p < .009$, $\eta_p^2 = .376$ (as well as a ninth order effect, $F(1,15) = 10.48$, $p < .006$, $\eta_p^2 = .411$, with no obvious interpretation). All aftereffects were significantly greater than 0 (all ts > 4.36 $ps < .001$), except at the 0% adaptor level ($t = 1.71$ $p = .109$), where no aftereffect was expected, and at the weakest adaptor level (40%) ($t = 1.07$ $p = .303$).

Angry/happy aftereffects. Inspection of Figure 2 shows that these aftereffects increased numerically up to a peak at 120%, which was close to the natural range boundary (M = 123). Over the natural range (0-120), polynomial contrasts revealed only a significant linear effect, $F(1,16) = 18.15$, $p < .001$, $\eta_p^2 = .531$. In addition, planned t-tests showed significant increases from 0 to 40 ($p = .012$), and 40 to 80 ($p = .027$), with no significant increase from 80-120 ($p = .618$). These results support opponent coding of expression.

Beyond the natural range, aftereffects remained high, with significant decreases only at 280 and beyond ($ps < .011$, uncorrected), which did not survive Bonferroni correction (Figure 2). Again, these results suggest that extremely distorted faces activate the expression coding system.

Over the full range (0-440), polynomial contrasts indicated significant quadratic, $F(1,16) = 17.47$, $p < .001$, $\eta_p^2 = .522$, and cubic effects, $F(1,16) = 11.60$, $p < .004$, $\eta_p^2 = .420$. All aftereffects were significantly greater than 0 (all ts > 2.32 $ps <$

.034), except those for 0% adaptors ($t = 0.30$ $p = .768$), where no aftereffect was expected, and 400% adaptors ($t = 2.09$, $p = .053$).

Summary. Expression aftereffects showed a significant linear increase over the full natural range for both expression–pair conditions, as predicted by opponent coding, and as found for identity aftereffects. Beyond the natural range, substantial aftereffects remained for highly distorted faces, also as found for identity aftereffects. For disgust/sad pairs, the aftereffects remained high right across the range. Inspection of Figure 1 suggests that even the most extremely distorted faces on the disgust and sad continua remained expressive, consistent with the robust expression aftereffects generated. For the angry/happy pairs, the aftereffects declined (although not significantly with Bonferroni correction) for the most extreme adaptors. Inspection of Figure 1 suggests a possible reason: the most extreme faces on the happy continuum are so distorted that they violate the face configuration, which would make them less effective adaptors for any holistic face representations. In addition, the absence of a mouth would eliminate any contribution of mouth-related adaptation.

Similarity Ratings

One participant, who appeared to have used the scale backwards, was excluded. For the other participants, similarity ratings were converted to dissimilarity scores (by subtracting scores from 11). Mean scores are plotted on Figure 3 for each of the two expression-pair aftereffect conditions. As expected, dissimilarity increased with morph distance from the average test face, although a flattening at very extreme (morph) levels suggests a ceiling effect. These observations were confirmed by a two-way repeated-measures ANOVA, with adaptor extremity as a repeated-measures factor and expression pair as a between-participants factor. There was a significant main effect of adaptor extremity, $F(4.32, 129.68) = 577.76$, $p < .0001$, $\eta_p^2 = .951$, with

significant linear, $F(1,30) = 2096.84$, $p < .0001$, $\eta_p^2 = .986$, quadratic, $F(1,30) =$ 218.12, $p < .0001$, $\eta_p^2 = .879$, and fourth order, $F(1,30) = 4.25$, $p = .048$, $\eta_p^2 = .124$, effects (all other Fs < 2.36, ps $> .135$). There was no main effect of expression-pair, $F(1,30) = 0.46$ $p = .502$, $\eta_p^2 = .015$, and no interaction, $F(4.32,129.68) = 0.99$ $p = .414$, $\eta_p^2 = .032$. Overall, these results confirm that our morphing procedure increased perceived adaptor extremity (dissimilarity from the average), albeit with a possible ceiling effect for impossibly distorted adaptors.
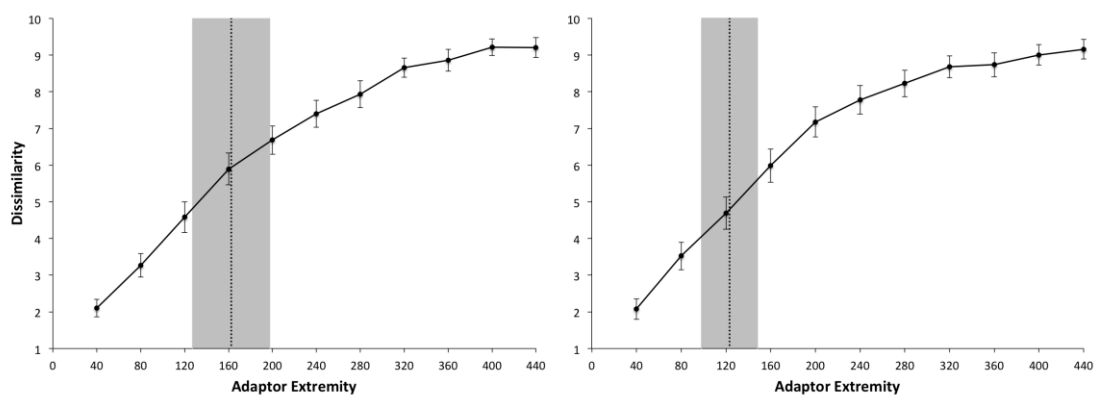

Figure 3. Mean dissimilarity of adaptor anti-expressions to the average expression for each expression-pair condition: Left: disgust/sad. Right: angry/happy. SE bars are shown.

Aftereffects as a function of *perceived* adaptor extremity

We show the aftereffects as a function of *perceived* adaptor extremity in Figure 4, which was created by rescaling the x-axis based on the dissimilarity ratings displayed in Figure 3. Note that this rescaling does not affect the monotonic increase in aftereffects observed over the natural range, which supports opponent coding. Nor can it alter the fact that pairwise differences beyond the natural range are not statistically significant (even though the numerical decline for the angry/happy continuum may appear steeper than before). Thus, these results do not alter any of our conclusions.
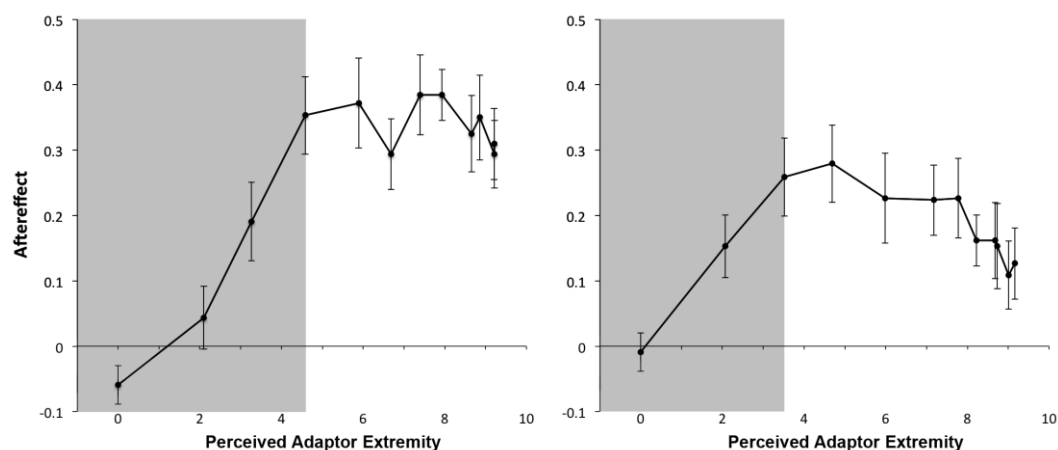
Figure 4: Expression aftereffects as a function of perceived adaptor extremity (dissimilarity from average expression) for disgust/sad (left) and angry/happy expression pairs (right). SE bars are shown. The shaded grey area to the left indicates adaptors perceived as being within the natural range of possible faces.

Discussion

We found that expression aftereffects increased linearly as adaptors became more extreme over the full natural range of possible faces. By explicitly determining (and spanning) the extent of that range, we can be confident that there is no decrease in aftereffects within that range that could indicate narrowband, multichannel coding. Instead, the results support norm-based, opponent coding of expression-related dimensions.

Converging evidence for opponent coding of expression comes from a different paradigm, in which participants adapt to either a central (average) expression or alternating expressions from opposite ends of an expression trajectory (e.g., anti-happy and happy) (Burton, Jeffery, Calder, & Rhodes, 2015). Specifically, the results rule out a three-channel model, with an additional, central channel tuned to the average. That model predicts opposite shifts in the two conditions. Adaptation of the central channel (by viewing an average expression) should selectively reduce sensitivity to the average expression, thus narrowing the range of expressions perceived as average. In contrast, adaptation to expressions from the opposite ends of a trajectory should selectively reduce sensitivity to those expressions, thus broadening the range perceived as average. However, the range of expressions perceived as central (average) actually narrowed in both conditions (Burton et al., 2015). This result is consistent with opponent coding, because both adapting conditions affect the two channels similarly.

Additional converging evidence comes from neurophysiological data. Face-selective cells in macaque monkeys show ramped response functions to many face features, consistent with opponent coding, rather than narrowly tuned responses that would suggest multichannel coding (Freiwald, Tsao, & Livingstone, 2009). Although

the focus of the study was on the coding of identity, some of the features, such as eyebrow slant, eye size and iris size, would certainly be relevant to coding expressions.

A second interesting feature of our results is that impossibly distorted faces generated substantial expression aftereffects. The same pattern has been found for identity (McKone et al., 2014, 2015; Robbins et al., 2007; Susilo et al., 2010). These findings indicate that highly distinctive/caricatured faces powerfully activate face-coding mechanisms, and may explain the effectiveness of grotesque faces as communicative devices in a range of artistic and other media. Strong aftereffects for impossibly extreme adaptors seem difficult to reconcile with narrowband, multichannel coding, because adaptation to extreme adaptors should have little impact on responses to the average test faces used here. More generally, it seems implausible that an efficient coding system would develop channels that are narrowly tuned to impossible, and thus rarely- or never-seen, configurations.

A third important feature of our results is that the entire complex pattern of aftereffects that increase over the full natural range and remain strong well beyond that range, mirrors the pattern seen for identity aftereffects (e.g., McKone et al., 2014, 2015). This parallel is consistent with a shared perceptual representation for expression and identity and is expected if there are common dimensions that contribute to the coding of both attributes. Critically, the results are similar within the natural range of possible faces, which is the relevant range for distinguishing between opponent and narrowband multichannel coding and the range to which face-coding mechanisms have been tuned by experience. Beyond this range, it is difficult to make precise comparisons, because the extent of impossibility is not matched across attributes (or across expression conditions). Nevertheless, the results are broadly

similar, with no significant decrease for either expression or identity aftereffects. We did see a numerical decrease for aftereffects in the angry-happy condition which could potentially be significant with greater power. This decrease is likely due to the disappearance of the mouth, which is an important cue to happy expressions, in extreme anti-happy adaptors (Figure 1). Ultimately, aftereffects must decrease as adaptors become so unfacelike that they no longer activate face-coding mechanisms. We did, however, find face aftereffects when the mouth was not visible (in extreme anti-happy adaptors). This result suggests that the full first order facial configuration (eyes above nose above mouth) is not required to engage and adapt expression-coding mechanisms. This result is consistent with our ability to perceive some expression even when parts of the face are not visible (e.g., when parts are obscured by sunglasses or other objects).

We measured the perceived dissimilarity of adaptors to the average test face, to check that increases in morph level extremity increase perceived extremity of adaptors. There was a very strong linear increase, with a much smaller quadratic effect reflecting a slight flattening at the most extreme levels where responses were close to ceiling (Figure 3). Inspection of Figure 3 confirms that the increase was linear over the natural range, so rescaling to perceived extremity would not change the shape of function relating aftereffect size to adaptor extremity in this critical range. The slight flattening that occurred beyond the natural range means that aftereffects in that range would reduce more rapidly when plotted as a function of perceptual rather than morph level units (see Supplementary Materials). However, rescaling would not change the fact that the decrease was not statistically significant, and to reiterate, a decrease in that part of the range can be consistent with either model.

We have interpreted our results as informative about higher-level expression coding mechanisms. We minimized the contribution of low-level, retinotopic adaptation, by using a size change between adapt and test faces and allowing free eye movements. We also minimized identity adaptation by using adapt and test faces of the same identity. A caveat is that the most extreme expression distortions might alter the apparent identity of the adapting faces (see Figure 1). Importantly, however, identity seems clearly preserved over the natural range, suggesting that we are not simply measuring identity coding over that critical part of the range. We cannot, however, rule out a contribution of mid-level shape adaptation, which raises the possibility that common coding of expression and identity arises at the level of shape coding mechanisms that feed into higher-level face-coding mechanisms.

There are certainly difficulties in using adaptation to explore visual coding mechanisms, particularly the lack of a one-to-one mapping between patterns of aftereffects and coding systems (Ross, Deroche, & Palmeri, 2014; Webster, 2015). For example, Ross et al (2014) have shown that both two-pool opponent coding (norm-based) and exemplar (cf. non-norm-based, multichannel) models can account for a range of aftereffect results, although they did not consider the precise pattern of increase across the full natural range seen here. They also found that a given model can generate a variety of outcomes, depending on the number of dimensions assumed in a face-space. This result raises the question of whether predictions derived from single dimensions, as done here, will "scale up" to more complex multi-dimensional face spaces. This question remains open.

Perhaps the core insight provided by tasks such as ours is about channel bandwidth. The monotonic increase in aftereffects over the natural range seen here is consistent with broadly tuned channels that encompass that full range, as with

opponent coding. In principle, however, it could also be accounted for by a multichannel model with broadly tuned channels, which is exactly what Ross et al's (2014) exemplar model is when its bandwidth is large. However, such a model would necessarily contain channels with large amounts of overlap, which would be massively redundant, with highly correlated activity between channels. It is also difficult to reconcile with our finding that high levels of adaptation are maintained across impossibly distorted adaptors. As noted above, to account for this finding, a multichannel system would require channels tuned to impossible and rarely- or never-seen distortions, but how could these ever develop? Finally, as outlined above, converging evidence from a different adaptation paradigm is consistent with two (opponent), but not three, broadly tuned channels (Burton et al., 2015). Overall, we suggest that there is a better case for opponent coding than multichannel coding of expression.

Concerns have arisen recently about the extent to which aftereffects, including face aftereffects, might result from changes in decision biases that reflect cognitive processes, in which case they would not be informative about perceptual representation (Morgan, 2014; Storrs, 2015). However, both bias changes and top-down effects more generally, can arise at any stage within a perceptual system that passes information from one level to another (Fodor, 1983; John-Saaltink, Kok, Lau, & de Lange, 2016; Teufel & Nanay, 2016). We suggest that, taken together, the striking changes in subjective experience, the tight links to adaptation and test durations, (Leopold, Rhodes, Muller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007), the relatively early temporal locus (Burkhardt et al., 2010) and the improved discrimination of identity following adaptation (Rhodes, Watson, Jeffery, & Clifford, 2010), make it unlikely that face aftereffects are purely cognitive effects.

Clearly, there are many challenges in using aftereffects to explore neural coding mechanisms. However, aftereffects yield unique information about the functional ways in which coding mechanisms adapt to changes in the visual input (Webster, 2015). In the context of the present study, we can say that expression-coding mechanisms adaptively recalibrate in very similar ways to identity-coding mechanisms.

In summary, we found that expression aftereffects increase over the full natural range of possible faces. There was no decrease within this range that would indicate narrowband, multichannel coding. Moreover, substantial aftereffects remained for impossibly distorted adaptors. This pattern is consistent with norm-based, opponent coding of expression-related dimensions. It also corresponds closely with that found for identity aftereffects (McKone et al., 2015), as expected if there are common dimensions that contribute to the high-level visual representation of both expression and identity (e.g., Calder & Young, 2005; Rhodes et al., 2015). Opponent coding highlights deviations from average values (signalled by equal activation of opponent pairs). We suggest that this kind of coding may help us perceive the subtle differences in facial appearance that underlie our expertise in discriminating expressions and identities.

References

Abboud, Hisham, Schultz, Hank, & Zeitlin, Vadim. (2008). Superlab 4.0 [Computer software]. San Pedro, CA: Cedrus Corporation.

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. *Journal of Vision, 9*(10), 10-10.

Afraz, S-R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision research, 48*(1), 42-54.

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science, 166*, 245-247.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Burkhardt, A., Blaha, L. M., Schneider Jurs, B., Rhodes, G., Jeffery, L., Wyatte, D., . . . Busey, T. (2010). Adaptation modulates the electrophysiological substrates of perceived facial distortion: Support for opponent coding. *Neuropsychologia, 48*(13), 3743-3756.

Burton, N., Jeffery, L., Calder, A. J., & Rhodes, G. (2015). How is facial expression coded? *Journal of Vision, 15*(1), 1-1.

Burton, N., Jeffery, L., Skinner, A. L., Benton, C. P., & Rhodes, G. (2013). Nine-year-old children use norm-based coding to visually represent facial expression. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1261.

Calder, A. J. (2011). Does facial identity and facial expression recognition involve separate visual routes? In A. J. Calder, G. Rhodes, M. H. Johnson & J. V. Haxby (Eds.), The Oxford Handbook of Face Perception (pp. 427-448). Oxford: Oxford University Press.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research, 41*(495-518).

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*, 641-651.

Clifford, C. W. G., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 267*, 1705-1710.

Fiorentini, C., Gray, L., Rhodes, G., Jeffery, L., & Pellicano, E. (2012). Reduced face identity aftereffects in relatives of children with autism. *Neuropsychologia, 50*(12), 2926-2932.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*: MIT press.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *nature neuroscience, 12*(9), 1187-1196.

Haxby, J. V., & Gobbini, M. I. (2011). Distributed neural systems for face perception. In A. J. Calder, G. Rhodes, M. H. Johnson & J. V. Haxby (Eds.), The Oxford Handbook of Face Perception (pp. 93-110). Oxford: Oxford University Press.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223-233.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*, 59-67.

Hoffman, E.A., & Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *nature neuroscience, 3*, 80-84.

Jeffery, L., Read, A., & Rhodes, G. (2013). Four year-olds use norm-based coding for face identity. *Cognition, 127*(2), 258-263.

Jeffery, L., Rhodes, G., McKone, E., Pellicano, E., Crookes, K., & Taylor, E. (2011). Distinguishing norm-based from exemplar-based coding of identity in children: Evidence from face identity aftereffects. *Journal of Experimental Psychology: Human Perception and Performance, 37*(6), 1824-1840. doi: 10.1037/a0025643

John-Saaltink, Elexa St, Kok, Peter, Lau, Hakwan C, & de Lange, Floris P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *the Journal of Neuroscience, 36*(23), 6186-6192.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *nature neuroscience, 4*(1), 89-94.

Leopold, D. A., Rhodes, G., Muller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences, 272*, 897-904.

McKone, E., Jeffery, L., Boeing, A., Clifford, C. W. G., & Rhodes, G. (2014). Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces. *Vision research, in press*.

McKone, E., Jeffery, L., Boeing, A., Clifford, C. W. G., & Rhodes, G. (2015). Corrigendum to "Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces"[Vision Res. 98 (2014) 1–13]. *Vision research, 108*, 119-120.

Morgan, M. J. (2014). A bias-free measure of retinotopic tilt adaptation. *Journal of Vision, 14*(1), 7-7.

Pond, S., Kloth, N., McKone, E., Jeffery, L., Irons, J., & Rhodes, G. (2013). Aftereffects support opponent coding of face gender. *Journal of Vision, 13*(14), 16.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17*, 48-63.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*(4), 473-497.

Rhodes, G., Jeffery, L., Clifford, C. W. G., & Leopold, D. A. (2007). The timecourse of higher-level aftereffects. *Vision research, 47*, 2291-2296.

Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. J. Calder, G. Rhodes, M. H. Johnson & J. V. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 263-286). Oxford: Oxford University Press.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., . . . Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition, 142*, 123-137.

Rhodes, G., Watson, T. L., Jeffery, L., & Clifford, C. W. G. (2010). Perceptual adaptation helps us identify faces. *Vision research, 50*, 963-968.

Robbins, R. A., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adapter position effects and neural models.

Journal of Experimental Psychology: Human Perception and Performance, 33(3), 570-592.

Ross, D. A., Deroche, M., & Palmeri, T. J. (2014). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review, 21*(1), 47-70.

Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science, 21*(9), 1248-1253.

Skinner, A. L., & Benton, C. P. (2012). The expressions of strangers: Our identity-independent representation of facial expression. *Journal of Vision, 12*(2), 1-13.

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology, 6*, 157.

Storrs, K. R., & Arnold, D. H. (2015). Face aftereffects involve local repulsion, not renormalization. *Journal of Vision, 15*(8), 1-1.

Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision research, 50*(3), 300-314.

Teufel, C., & Nanay, B. (2016). How to (and how not to) think about top-down influences on visual perception. *Consciousness and cognition*.

Valentine, T. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception, 15*(5), 525-535.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161-240.

Valentine, T. (Ed.). (1995). *Cognitive and Computational Aspects of Face Recognition: Explorations in face space*. London: Routledge.

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science, 1*, 547-567.

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society: Biological Sciences, 366*, 1702-1725.